ACCEPTANCE

This dissertation, STUDENT ACHIEVEMENT IN SCIENCE AND MATHEMATICS IN URBAN PROFESSIONAL DEVELOPMENT SCHOOLS DURING FIRST YEAR OF IMPLEMENTATION, by SUSAN L. OGLETREE, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree Doctor of Philosophy in the College of Education, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chair, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty. The Dean of the College of Education concurs.

____
William L. Curlette, Ph.D.
Committee Chair

____
Mary P. Deming, Ph.D.
Committee Member

____
Gwendolyn Benson, Ph.D.
Committee Member

____
Roy M. Kern, Ed.D.
Committee Member

____
Douglas Davis, Ph. D.
Committee Member

____
Date

____
Sheryl A. Gowen, Ph.D.
Chair, Department of Educational Policy Studies

____
Ronald P. Colarusso, Ed.D.
Dean, College of Education

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's director of graduate studies and research, or by me. Such quoting, copying or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation, which involves potential financial gain, will not be allowed without my written permission.

____
Susan L. Ogletree

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Susan L. Ogletree
P.O. Box 1627
Mableton, GA 30126

The director of this dissertation is:

Dr. William L. Curlette
Department of Educational Policy Studies
College of Education
Georgia State University
Atlanta, GA 30303

VITA

Susan L. Ogletree

ADDRESS: P.O. Box 1627
Mableton, GA 30126

EDUCATION:

Ph.D.	2007	Georgia State University Educational Policy Studies
Ed. S.	2004	Georgia State University Professional Counseling
M.S.	2002	Georgia State University Professional Counseling
Ed.S.	1985	West Georgia College Educational Leadership
M.Ed.	1974	West Georgia college Elementary Education/Educational Leadership
B.M.	1973	West Georgia College Music

PROFESSIONAL EXPERIENCE:

2004 – present	Director, Educational Research Bureau Georgia State University
2003 – 2004	Grants and Contracts Officer, Educational Research Bureau Georgia State University
2002 – 2003	Administrative Assistant, Educational Policy Studies Dept Georgia State University
1989 – 2001	Founding Head of School The Georgian School, Carrollton, GA
1975 – 1989	Head of School Oak Mountain Academy, Carrollton, GA
1973 – 1975	Band Director/Assistant Principal Carroll County Schools, Carrollton, GA

PUBLICATIONS:

Sprouse, D., Ogletree, S., Miller, M., Granville, H., & Kern, R. (2005). An Adlerian model for alliance building. *The Journal of Individual Psychology, 61*(2), 136-148.

PRESENTATIONS:

Benson, G., Curlette, W. L., Ogletree, S. L., & Taylor, D. (2006, April 6-11). *Research and evaluation in the Professional Development School Programs Deliver Success grant*. Paper presented at the annual meeting of the American Education Research Association, San Francisco.

Benson, G., Curlette, W. L., Ogletree, S. L., & Taylor, D. (2006, March). *RA in PDS: Research and assessment for revealing advancement in the Professional Development School Programs Delivers Success grant.* Paper presented at the Professional Development Schools National Conference.

Benson, G., Curlette, W. L., Ogletree, S. L., & Taylor, D. (2006, January). *PDS: How to stand and deliver*. Paper presented at the meeting of the American Association of Colleges for Teacher Education.

Curlette, W. L., & Ogletree, S. L. (2006, October). *Using dreams and early recollections in counseling practice.* Paper presented at the South Carolina Conference of Adlerian Psychology.

Curlette, W. L., & Ogletree, S. L. (2003, October). *Interpreting BASIS-A Inventory results: Combining theory, research and practice.* Presentation at the annual meeting of the South Carolina Adlerian Society, Myrtle Beach, SC.

Kern, R. M., Ogletree, S. L., & Miller, M. (2004, February). *Adherence, therapeutic alliance and Individual Psychology: The mystery of client no shows.* Paper presented at the Florida Adlerian Society 2004 Annual Conference of Individual Psychology.

Ogletree, S. L., & Kern, R. M. (2003, February). *Brief therapy is more than one session.* Paper presented at the Florida Adlerian Society 2003 Annual Conference of Individual Psychology.

Ogletree, S. L., Kern, R. M., & Miller, M. (2002, October). *You can't help them if they're not there.* Paper presented at the South Carolina Conference of Adlerian Psychology.

ABSTRACT

STUDENT ACHIEVEMENT IN SCIENCE AND MATHEMATICS IN URBAN PROFESSIONAL DEVELOPMENT SCHOOLS DURING FIRST YEAR OF IMPLEMENTATION

by

Susan L. Ogletree

Using a quasi-experimental design, the author examined the effects of the Professional Development School Partnerships Deliver Success educational model on student academic achievement in science and mathematics in 12 high-needs, urban elementary, middle, and high schools in the southeastern United States. Student achievement was measured for first to eighth grade students by the State Criterion-Referenced Competency Test and for 11th-grade students by the State High School Graduation Test. 6 ANOVAs were used to compare baseline and year 1 performance data. Student ethnicity was used to disaggregate the data to investigate the extent, if any, to which achievement gaps narrowed. For the different ethnic groups, the small changes in proportion passing across the first year of implementation were not correlated with mean scale score changes as measured by Hedges's g effect sizes. This result has national implications for the *No Child Left Behind Act of 2001* policy in terms of reporting results. Three of the 6 ANOVAs showed significant change in achievement means for the PDS schools when using PDS school data only. However, when data from both PDS and matched comparison schools were analyzed, the overall results indicated no statistically significant gains in mathematics and science means for the professional development schools in

relation to the comparison schools for the first year of professional development school implementation.

STUDENT ACHIEVEMENT IN SCIENCE AND MATHEMATICS IN URBAN PROFESSIONAL DEVELOPMENT SCHOOLS DURING FIRST YEAR OF IMPLEMENTATION

by

Susan L. Ogletree

A Dissertation

Presented in Partial Fulfillment of Requirements for the
Degree of
Doctor of Philosophy
in
Educational Policy Studies
in
the Department of Educational Policy Studies
in
the College of Education
Georgia State University

Atlanta, Georgia
2007

ACKNOWLEDGMENTS

The path toward completion of my doctorate has been enlightening both personally and educationally. Because of this exceptional opportunity, I have come to know and be supported by many people. The first person I would like to acknowledge is Dr. William L. Curlette, Chair of my Dissertation Committee. His inspirational teaching, experienced advice and continuous support throughout the dissertation process were exceptional. I would also like to thank the members of my committee, Drs. Gwendolyn Benson, Mary Deming, Roy Kern, and Douglas Davis for their exceptional support and encouragement.

I extend thanks to friends and colleagues Dr. Sheryl Gowen, Sherry Kirby, Dr. Susan McClendon, Harley Granville, Sujatha Bhagavati, Shane Blasko, and Kevin Barnaba for their encouragement. Special thanks goes to Wm S Boozer, Dr. Dee Taylor, Don Segal, and Alison Norsworthy for their continuous advice and support in completion of my dissertation.

I owe a special note of gratitude to my family, August Ogletree Dale, Tee Ogletree, and Laurie Forstner. They have been actively involved in my academic studies over the years as only family members can be. I have always and will continue to appreciate their encouragement and enthusiasm for my chosen paths.

Finally, I extend a special thanks to Robert Hendrick for his patience, knowledge and support in helping me to achieve my goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AACTE	American Association of Colleges for Teacher Education
AYP	Adequate Yearly Progress
CRCT	Criterion-Referenced Competency Tests
GHSGT	Georgia High School Graduation Tests
GPS	Georgia Performance Standards
MSDE	Maryland State Department of Education
MSPAP	Maryland School Performance Assessment Program
NCATE	National Council for Accreditation of Teacher Education
NCLB	No Child Left Behind Act of 2001
NCREST	National Center for Restructuring Education, Schools, and Teaching
NNER	National Network for Educational Renewal
PDS	Professional Development School
QCC	Quality Core Curriculum
SAT-9	Stanford Achievement Test-9
SES	Socioeconomic Status
TAAS	Texas Assessment of Academic Skills
WVU	West Virginia University

CHAPTER 1

INTRODUCTION

Since the publication of *A Nation at Risk* in 1983, many additional reports have been published criticizing the United States education system and the institutions that produce teachers. In 1986, the publication of *A Nation Prepared*, the report of the Carnegie Forum on Education and the Economy, and the Holmes Group trilogy, *Tomorrow's Teachers* (1986), *Tomorrow's Schools* (1990), and *Tomorrow's Schools of Education* (1995), called for significant restructuring of U.S. teacher education programs and influenced the birth of the Professional Development School movement in the United States (Campoy, 2000; Stallings & Kowalski, 1990). These reports went beyond criticism and for the first time offered solutions to the growing U.S. education problem; they managed to capture the attention of both educational leaders and policymakers throughout the United States. The solutions brought forth in both publications set the terms for the continued debate surrounding teacher preparation in this country (Fraser, 1992).

A professional development school (PDS) is a formal, long-term partnership among a school district, one or more K-12 schools, and a university. The partnership is established to share the responsibility for the preparation of beginning teachers, the further professional development of experienced teachers, and the improvement of practice with the overarching goal of improving student achievement (Levin, 2002). The PDS model encourages teacher educators to use field-based clinical preparation of novice teachers (Darling-Hammond, 1996; Goodlad, 1984, 1990; Holmes Group, 1986, 1990, 1995). While seeking to improve student achievement, PDSs also provide opportunities for professional development, teacher empowerment, student diversity, and equity in education – most requiring significant systemic change.

The implementation of PDSs is one way that teacher education institutions are answering the call for reform. The movement is gaining momentum, and it appears to be more than a passing fancy (Rice, 2002). Robinson and Darling-Hammond (1994) proclaimed that PDSs are so important to reform that they are becoming required. Educators who work to implement PDSs agree on two major points. The first is that they are costly in time and labor. The second is that when a PDS functions well, it improves the teaching methods of experienced K-12 teachers, teacher educators, and credential candidates (Robinson & Darling-Hammond, 1994; Zimpher, 1990).

Many researchers have tried to document the effectiveness of PDSs. Their studies primarily addressed the theory, implementation, and description of PDSs, while others have explored the nature and impact of district-school-university partnerships (Abdal-Haqq, 1988; Book, 1996; Compoy, 2000). Most current research is qualitative with little methodological detail. The lack of detail calls into question the validity, reliability, and replicability of the studies (Book, 1996). Stallings and Kowalski (1990) reported that while the number of PDSs has grown, there is little systematic evaluation of them. Eleven years later, Reed et al (2001) continued to bemoan the lack of systematic evaluation of PDSs.

In 2000, Teitel emphasized that in order for the PDS movement to continue to develop in the United States, systematic research must occur. While acknowledging the inherent difficulties of evaluating a multifaceted program, Teitel pointed out that PDSs must engage in quality research that is carefully constructed and implemented in order to show its effectiveness. Abdal-Haqq (1996) specifically stated that there was a need for more research on the effect of PDSs as it relates to student academic achievement. It has become clear that without systematic student achievement data that show PDSs make a difference in student academic achievement, the PDS movement could wither and die.

Teitel (2000) discussed the factors that make systematic evaluation of PDSs problematic. A few of the most significant factors are the fragility of collaborative partnerships, the difficulty of qualitative research including comparison groups, disagreement among stakeholders on the importance of outcomes and how to measure them, and misalignment of program goals and construction of the evaluation. Given these research difficulties, it comes as no surprise that there are few quantitative systematic assessments of student achievement over time available to PDS researchers and stakeholders.

Research Questions

Using quantitative analysis (Cooper & Hedges, 1994; Stevens, 2002), baseline and year-1 implementation data of a 5-year Teacher Quality Enhancement Grant was compared. The grant known as the Professional Development Schools Deliver Success "PDS^2" intervention is based on individualized strategic plans collaboratively designed by school faculty, a university coordinator, and the design team. All schools participated in a strategic planning retreat where school faculty, the university coordinator, and the design team collaboratively developed an individualized school strategic plan. The design team consisted of the project investigator, the director of research, the project director, the budget director, and one university coordinator each from the university's Early Childhood Department and Middle/Secondary Department depending upon the grade level of the participating school. Each PDS school received a university coordinator (funded by the grant) to work one day per week in schools, in addition to facilitating preinterns and interns who were placed in the school. The grant funds were provided to purchase materials and supplies for academic program enhancement and encouragement of action research projects. Additional funding was given to each PDS school for professional development needs as identified by the school. PDS school coordinators received a stipend to support and coordinate the data collection in each school and teachers that participated in the data collection process also received a stipend. Additionally, a part-time data manager was written into the grant for each participating school system. This person, employed by the school system, complied data for the Director of Research as required by the federal government for the Teacher Quality Enhancement grant. Each participating school system received significant funding for support of professional development school activities.

I examined the effects of the federal grant Professional Development School Partnerships Deliver Success also known as "PDS^2" on three randomly selected feeder-patterned, high-needs urban schools from four different school systems. I examined the following questions:

1. How does the PDS^2 model affect mean student achievement in mathematics and science as measured by the Georgia Criterion Referenced Competency Test and Georgia High School Graduation Test?

2. Are there significant differences in mean achievement test scores between PDS^2 feeder pattern schools and comparison schools?
3. From the baseline year to the end of the first year, how many PDS_2 and Comparison Schools have changed their Adequate Yearly Progress status and in what direction?
4. Is there a mean difference between ethnic groups on the scaled scores of the CRCT and HSGT for mathematics and science?
5. From the baseline year CRCT and HSGT to the end of the first year across ethnic groups, is there a mean difference in scaled scores on the CRCT and HSGT for mathematics or science?
6. From the baseline year CRCT and HSGT to the end of the first year for different ethnic groups, is there a correlation between the proportion passing change on the CRCT and HSGT for mathematics or science and Hedges's g effect size?

Purpose

I examined and compared the effects of PDS^2's educational activities on student academic outcomes in mathematics and science for twelve high-needs urban schools using students' scores on the Georgia Criterion-Referenced Competency Tests (CRCT) and the Georgia High School Graduation Tests (GHSGT). Mathematics and science were chosen because the state is currently changing the test to reflect a new curriculum and these two areas had the highest number of students taking the CRCT, and were the least affected by the curriculum test change. The study provides PDS impact information based on CRCT scores for educators and stakeholders. It adds to the quantitative body of knowledge as it relates to academic achievement in professional development schools and contributes statistical information for the overarching research question of the effects of PDS school participation on student academic achievement.

Significance

Considering the lack of PDS research on student achievement, this study has made a significant contribution to PDS academic achievement research when looking at baseline and Year 1 implementation CRCT test scores. Moreover, because of the analysis and comparison of the CRCT test scores of twelve urban feeder pattern schools and twelve comparison schools, this study is an accountability measure of a particular PDS implementation on student academic achievement. The study also documented the educational program implementations used in the PDSs during the baseline and year 1 data collection period. Finally, the PDS initial implementation documentation and analysis of CRCT and GHSGT scores provided a comprehensive model of measuring and determining the effects of PDSs on academic achievement.

Definitions

The abbreviation for Criterion-Referenced Competency Tests is CRCT. Georgia law, as amended by the A+ Education Reform Act of 2000, requires all students in grades 1 through 8 to take the CRCT in reading, English/language arts, and mathematics. Students in grades 3 through 8 take tests in science and in social studies in addition to the three areas previously mentioned. The CRCT is aligned to the Quality Core Curriculum (QCC), which has been used to gauge the quality of education throughout the state of Georgia (Georgia Department of Education, n.d.a.).

In 1985, the Quality Basic Education Act was passed and required Georgia to maintain a curriculum that specifies what students are expected to know in each subject and grade. The Quality Core Curriculum was developed from this legislation, and it is a guideline for instruction that helps teacher, students, and parents know what topics are to be covered and mastered in each course. The QCC states minimum requirements for each course to be taught.

The abbreviation for Georgia Performance Standards is GPS and is an update of the QCC standards. In January 2002, a *Phi Delta Kappa* audit reported that the QCC lacked depth, could not be covered in a reasonable amount of time, and did not meet national standards (Georgia Department of Education, 2005). The GPS is the revised and strengthened curriculum that currently drives instruction and assessment in the state of Georgia. The CRCT will now be aligned with the new Georgia Performance Standards (Georgia Department of Education).

The Georgia High School Graduation Tests were mandated by the 1991 Georgia law O.C.G.A. section 20-2-281. This law requires that a curriculum-based assessment be administered in the 11th grade for graduation purposes. Students in the state of Georgia who seek a high school diploma must pass all five tests on the HSGT, which cover English/language arts, writing, mathematics, science, and social studies. Students first take the graduation test in their junior year, with the writing test being given in the fall and the other four tests administered in the spring (Georgia Department of Education, n.d.b.).

Adequate Yearly Progress (AYP) is a series of performance goals that every school as a whole must achieve within times frames established by the *No Child Left*

Behind Act of 2001. To make AYP, 95% of students enrolled in each group must participate in the AYP assessments and each group of students must meet or exceed established statewide annual objectives (GreatSchools Inc., n.d.).

A professional development school (PDS) is a school based on the guiding principles established by the Holmes Group (1986, 1990, 1995). Currently, the National Council for Accreditation of Teacher Education (NCATE, 2001) defines a PDS as a school that was redesigned and restructured to meet better the continuous educational needs of both teachers and students. The overarching goal of a PDS is to increase student academic achievement through the use of the five NCATE standards. The five standards that address the characteristics of PDSs and are used to increase student achievement through improving teacher quality are learning community, accountability and quality assurance, collaboration, equity and diversity, and structures, resources and roles (NCATE, 2001). Educational activities designed to improve teacher quality such as the establishment of professional learning communities and needs driven professional development are provided in PDS schools. Collaborative action research projects developed to answer specific classroom research questions are encouraged and supported in PDS schools. Findings from the action research projects are published for collaborative teacher use. Teachers in PDS schools continuously review educational data collected in an effort to plan meaningful educational activities designed to increase academic achievement. Finally, administrators and teachers are encouraged to participate in conversations surrounding equity and diversity within the school and community setting.

Many different labels have been used to describe schools that are based on the PDS model. Examples of these labels are partner school, professional development academy, and induction school. Abdal-Haqq (1998) identified PDS as the most widely used and accepted label, and this is the name used in this research.

PDS^2 is the abbreviation for Professional Development Schools Partnerships Deliver Success. The two primary goals of this $6.1 million federal grant are to increase student achievement and to increase teacher retention across four metropolitan southeastern public schools systems. The partners are Georgia State University's Colleges of Education and Arts & Sciences, Atlanta Public Schools, DeKalb County Schools, Fulton County Schools, Gwinnett County Schools, Georgia Perimeter College, Clark Atlanta University, and the Georgia Association of Educators. After review by Institutional Review Boards, officials in each school system approved releasing the data to address federal grant requirements as related to student achievement.

Assumptions and Limitations

This first assumption of my study is that the CRCT and GHSGT are meaningful assessments of student achievement. The CRCT tests are designed to measure how well students acquire the skills presented in a specific curriculum or unit of instruction. The CRCT is intended to test Georgia's content standards as outlined in the QCC and GPS. Also, the GHSGT assesses a sample of the knowledge and skills acquired during a high school education. The knowledge and skills assessed on the GHSGT were selected by Georgia educators and curriculum specialists. The five test areas included on the GHSGT are social studies, English/language arts, mathematics, science, and writing. These tests are based on the standards specified in the QCC as established by the State Board of Education and revised in November 1997. The results of this study will be based on the assumption that these tests are valid and reliable.

The groups in an untreated control group design with separate pretest and posttest samples are not equivalent. However, for this study comparisons can still be made because the Comparison Schools were matched with the PDS schools on percentage of free and reduced lunch, ethnic group, and previous student achievement. Within a school system, if more than one feeder pattern qualified as a set of matched schools for the PDSs, then the comparison schools were randomly assigned. Comparisons among pretest subgroups are informative about equivalence of the posttest subgroups as well (Cook, Shadish & Campbell, 2002).

A third assumption is that students tend to remain at the same national percentile rank over time if there are no interventions. This is known as the equipercentile assumption (Tallmadge & Wood, 1981).

A fourth assumption is the quality of the data provided by the school systems. It is assumed that the data provided by the school systems met the requirements as requested by the researcher. The researcher requested that the data be reported for first time test takers when appropriate in the same way that data from the school systems is reported to the State of Georgia. It was requested that these data be reported for the baseline year (Spring 2005) and the end of the first year of the PDS^2 grant (Spring 2006). Data were requested for all students in these schools as reported to the Department of Education in the State of Georgia.

This study also has some limitations. The first and perhaps most significant limitation is the Georgia Department of Education's decision to make significant changes in the test items that appear on the CRCT. While the mathematics and science tests have not been changed from the baseline year to year 1 of data collection, the reading and language arts have been changed. Currently, an equating study has not been published; therefore, reading and language arts will not be included in this analysis of data. Also, mathematics and science scores using the new test will not be used.

The second limitation is the use of an untreated control group design with separate pretest and posttest samples. This design is the most frequently used in the social sciences. Drawing causal conclusions using this design is difficult because of threats to internal validity.

The third limitation is related to the use of student achievement as the dependent variable. Because identifying specific educational strands (independent variables) and activities related to them are difficult, identifying direct causation links from variable to outcome is virtually impossible. The complexity and interconnectedness of educational strands within a PDS may help to explain the current lack of empirical studies conducted in PDSs (Abdal-Haqq, 1998; Teitel, 2000, 2001). In spite of this, there is reason to conduct the research. The primary reason is to add to the empirical body of PDS knowledge that already exists so that when the collected PDS data are holistically analyzed it will show if the PDS^2 model improves or does not improve student achievement.

Summary

In this chapter, I have introduced a study of data collected over a 15-month period using an experimental, untreated control-group design. The data were collected for evaluation as part of a large federal grant studying PDSs and non-PDSs. The findings of this study add to the empirical body knowledge of research that has previously been conducted in PDSs. Additionally, the purpose of the study, research questions, assumptions, and limitations are discussed. The following chapter is the review of the literature and is presented in two sections. The first section focuses on the history and development of the PDS movement in the United States. The second section focuses on the impact of the PDS movement on student achievement.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter focuses on current research and evaluation that has been conducted concerning Professional Development School partnerships and academic achievement as it develops within a PDS environment. The first section of the literature review focuses on the history of the PDS movement in the United States and its subsequent development. The second section of the literature review focuses on the impact of the PDS movement on student achievement.

History of the Professional Development School Movement

The PDS movement has its origins in the late 19^{th} century work of John Dewey. By establishing the first laboratory school in Chicago, Dewey (1896) sought to approach teaching scientifically by testing, verifying and criticizing theoretical statements and principles related to teaching as well as by adding to the early body of knowledge related to the field of teaching. The first Dewey laboratory school was established with a focus on educational research in an effort to help document and improve teaching methods. The school was staffed with master teachers, who would practice and model the art and science of teaching. Beginning teachers spent time training in the laboratory school, which provided practice teaching experiences prior to entering the profession. This experience was analogous to the laboratory experience in the fields of biology or physics (Stallings & Kowalski, 1990). The laboratory school would be used to accumulate a body of knowledge to support educational practices through testing, verification, collection, and dissemination of data (Campoy, 2000). Goodlad (1980) reported five additional goals of the laboratory school in addition to its scientific mission: (a) use best practices to educate students, (b) develop and test new and innovative teaching methods, (c) encourage research and development, (d) prepare novice teachers, and (e) provide professional development services for experienced teachers.

During the early 20th century, efforts began to be made toward professionalization of many occupations. The legal, medical, and teaching occupations were among these. To move from an occupation to a profession required a commitment to extensive and standardized educational programs. This movement effectively excluded those individuals who were poorly trained to practice within the profession. In 1910, doctors could be trained in a myriad of unregulated programs ranging from three weeks of symptom memorization to multiple years at Johns Hopkins University, where students studied medicine as a science and participated in clinical internships at the university supported teaching hospital (Darling-Hammond, 2006). The initial focus on professional training for all occupations began with the Flexner Report on Medical Education in the United States and Canada (bulletin #4), published and released by the Carnegie Foundation for the Advancement of Teaching (1914).

Later in 1920, the Professional Preparation of Teachers for America Public Schools (bulletin #14) was published and became known as the “Learned Report.” The report borrowed the framework from the Flexner Report. The Learned Report recommended that teacher education become a professional, evidence-based clinical preparation program similar to that recommended by the Flexner Report for the practice of medicine (Imig & Imig, 2005; Levin, 1992). This recommendation would require the strengthening of teacher preparation in the United States and effectively end the practice of allowing untrained, individuals to teach (Darling-Hammond, 2006).

One of the most important recommendations made by the Learned Report was that all teachers be required to participate in a 4-year prescribed teaching curriculum. The idea was not only to become proficient in the subject matter but also to learn the art and science of teaching. Future teachers were expected to attend a college whose main focus was that of preparing future educators. Faculties at those educational institutions were expected to work with K-12 schools to produce teachers with full knowledge of the "scholarship of teaching." The same report encouraged the use of clinical based education programs so that teachers in training would have the opportunity to observe and practice before moving into their own classroom. It was also recommended that college faculty establish and maintain on-going practice relationships while frequently visiting local schools. The importance of bridging the gulf between the theoretical and the practical through these relationships was also included in the report. Learned envisioned that professional schools of education would be held in equal esteem as those that prepare doctors or lawyers. However, during the past 85 years, the Learned Report and its recommendations have disappeared from teacher education reform discourse (Imig & Imig, 2005).

Another precursor to the PDS was the portal school, which began in the late 1960s. The portal school concept was first introduced as the Florida State University Model for the Preparation of Elementary School Teachers, where a particular process of teacher education was advocated. Sowards (1969) described the portal school as a place for new and experienced teachers interested in the improvement of children's learning.

The portal school was also used by school districts to introduce new curricula and recommended best practices. Each school district established one portal school at elementary, middle, and high school levels. These schools were identified as places for interaction and collaboration between university and school faculty in order to identify necessary teaching competencies to support innovative teaching strategies that had been shown to be effective.

While each portal school was created to meet the needs of the community it served, there were four common elements:

1. Advisory councils to act as liaison between the participants and the building principal
2. Portal school selection made with support of all participants from the university and school system.
3. Planning time provided for developing projected goals for all participants from the university and the school system.
4. Formal agreements executed among schools, colleges, state departments of education, community, and teachers for administration, evaluation, and revision of educational programs. (Lutonsky, 1972, p. 8)

Located within each portal school were university and school system programs for preservice and inservice training and curriculum development. The university funded a full-time professor for each identified portal school to assist in the development and coordination of a variety of consultancies that included testing, professional development workshops, staffing, and program development. During the 1970s, portal schools were developed at Florida State University, Temple University in partnership with the Philadelphia Public Schools, and University of Georgia with the Atlanta Public Schools. By 1980, the term “portal school” had been dropped from the literature possibly because of the lack of quantitative research required to assess student outcomes (Stallings & Kowalski, 1990).

In 1983, with the publication of *A Nation at Risk* by the National Commission on Excellence in Education, the U.S. public was made aware of the plummeting achievement test scores of both lower achieving students and college-bound students. Based on standardized test results, both segments of the population were mastering less of the academic subject matter than in the previous decades. The U.S. educational performance was described as "a rising tide of mediocrity" that threatens the educational foundation of American society (*A Nation At Risk*, 1983). In his book, *A Place Called School*, Goodlad (1984) reported that U.S. schools were in trouble. As a direct result of these national publications, state and local commissions of excellence were created to examine the educational problems. These commissions recommended changes in school curriculum, graduation requirements, teacher certification, and assessment (Stallings & Kowalski, 1990).

Business and industry leaders were also concerned about the decline in education. They declared a need for employees who were proficient in basic skills and able to transfer knowledge. They declared that employees also needed to have a thorough understanding of information acquisition through the use of computers and expanded technology. Developing this type of employee and to restructure the schools for the 21st century would require a new paradigm, including construction of a new and different system of teacher education. The new paradigm would require different types of principals, teachers, and colleges of education (Levin, 1992; Stallings & Kowalski, 1990).

The Carnegie Forum on Education and Economy (1986) and the Holmes Group Report encouraged the reconceptulization of teacher education at the university level. The

Holmes Group report recommended that universities move toward a collaborative model of professional-school partnerships linking colleges of education with K-12 schools where inquiry and action come together in reflective practice (Levin, 1992) while the initial recommendation by the Carnegie Forum on Education and Economy was the establishment of a National Board for Professional Teaching Standards. The Holmes Group recommendation was initiated in an effort to increase the content knowledge of teachers in the field. The Carnegie forum recommendation was made in an effort to enhance the prestige of the teaching profession through recruitment of more academically able candidates. Toward that end, the Carnegie Forum also recommended the establishment of clinical schools. These schools would serve to prepare teachers to meet the new standards and link colleges of arts and sciences faculties and college of education faculties with K-12. The clinical schools would provide K-12 teachers with collaborative access to college faculty with extensive content and pedagogical knowledge.

The Carnegie Forum on Education and Economy (1986) recommended that professional teacher training be taught at the graduate level establishing two-year training for teachers. This model was comparable to the teaching hospital model. Exemplary public school teachers would hold adjunct appointments at the university and teach in the Master's-degree program, allowing for rich collaborative dialogue among them. University and public school teachers would be given time to reflect and engage in discourse on teaching and learning practices within the clinical school environment. Teacher candidates would complete a two-year graduate course of study with coursework and internship occurring within the first year and residency with direct supervision occurring within the second year (Stallings & Kowalski, 1990). In effect, teachers would no longer be trained during the undergraduate years. Teacher training would occur at the graduate school level.

The Holmes Group (1986, 1990, 1995) is often credited as the first to conceptualize and give the initial boost to the idea of professional development schools. While this is certainly the case, the roots of the movement can also be found in school-university partnerships, reform of teacher education, the professionalization of teaching, and the use of standards as a developmental framework. In addition to the Holmes Group, many national organizations and smaller initiatives supported the PDS movement. The PDS movement received national support from the American Association of Colleges for Teacher Education (AACTE), the National Council for Accreditation of Teacher Education (NCATE), and the National Center for Restructuring Education, Schools, and Teaching (NCREST; Teitel, 1999).

Professional development schools were originally conceived as teaching communities housed in regular schools that would connect colleges of education with K-12 practitioners. The school-university collaborative would seek to develop excellent learning programs, thought-provoking teacher preparation, professional development for all participants, and research projects to enhance pedagogical knowledge (Campoy, 2000). These ideas were similar to the ones expressed in the Carnegie Forum on Education and Economy (1986). The collaborative partnership also sought to bring together teacher candidates, practicing teachers, administrators, and university faculty for rich, engaging discourse. Goodlad's National Network for Educational Renewal (NNER), NCREST, and the NCATE agreed upon four basic goals that PDSs must accomplish.

These goals were

1. Provide a clinical setting for pre-service education
2. Engage in professional development for practitioners
3. Promote and conduct inquiry that advances knowledge of teaching
4. Provide exemplary education for P-12 students (Clark, 1999; Teitel, 1998)

Teitel (1998) reported that collaboration, reform, and renewal are the three most important strands used to develop the PDS movement. Historically, teaching in a classroom had been a highly isolating and challenging activity. The culture of public schools encouraged separation and low interdependence among teachers. Teachers would close their doors in egg-crate schools knowing that what they did in their own classroom did not affect other teachers' work (Cohen, 1981). It was not until the 1980s that collaboration among faculty members became an important part of educational reform. The next logical step in collaboration was creating partnerships across institutions, such as between school and university. Establishing a partnership of this kind engaged complementary expertise and created a working synergy allowing both partners to expand their educational boundaries while challenging accepted educational practices that are the basis for teacher decision making in the classroom (Pugach & Johnson, 2002). Work within the PDS movement has confirmed that there is also a synergy between collaboration and learning (Neapolitan & Scott, 2004). Continuing to find more resources to support PDS work in the areas of time, space, people, and money is necessary in order to sustain collaborative synergy (NCTAF, 1996). Collaborative partnerships are relational and inclusive in nature. The inclusivity increases energy and joint ownership in the partnership. The established relationship helps to decrease the tendency to find fault for perceived problems while increasing the tendency to work together to solve problems (Pugach & Johnson, 2002).

Collaborative partnership building between school and university can also be used to create an environment for resolving tensions historically existing between them (Sewall, Shapiro, Ducett, & Stanford, 1995). One historically documented disconnect between school and university is the theory-to-practice dichotomy (Stoddard, 1993). Preservice teachers often report being frustrated to find classroom settings dissimilar to those studied in educational methods courses. It takes the combined knowledge of theory and practice to help create a teacher who can be successful in the reality of classrooms today. When rich collaboration occurs between universities and schools throughout the teaching process, the resulting new approaches have greater impact and use in the classroom. The conflict between theory and practice can be used to encourage a synergistic relationship between school and university promoting exploratory discussion so that mutually acceptable approaches to solving the problem can be found (Wiseman & Cooner, 1996). More importantly, the partnership allows the student teacher exposure to the intersection between theory and practice in an authentic classroom setting. The student teacher will also experience changes as they occur in the teacher education program (Su, 1999). While the model for teacher education is still seen as theory to practice, moving theory to practice exploration from the university into the classroom is imperative. This approach rejects the notion that accumulating content knowledge automatically qualifies one to teach (Campoy, 2000).

Clark (2000) argued that one purpose of a PDS is to support a common philosophical foundation for the efforts of school improvement. To gain consensus for a common PDS philosophy, the university and school must spend collaborative time in open discourse examining and exploring many different points of views is necessary.

School faculties, pressured to find quick-fix solutions to complex, multi-layered problems, often state that there is no time for this type of academic, theoretical activity. University faculty call for a more reflexive approach to problem solving that can lead to the discovery of more long-lasting, satisfactory solutions. The differences in the problem-solving approaches illustrate the tension created between a need to react quickly and more long-term reflexive thinking that require data collection and discussion. Different cultures result in different kinds of behavior. Despite these differences, the rewards of school-university collaboration can help broaden the perspective about what is institutionally possible (Trubowitz & Longo, 1997).

Cross-institutional collaborative partnerships are in a unique position to support real and lasting change in schools and universities. The opportunity to explore the problems of educational practice together, identifying, researching and planning for change maximizes the potential of better meeting the needs of a changing diverse student population at both the K-12 and university level. Snyder (1999) suggested that educational change is particularly difficult when it involves the politics, philosophy, and implementation of new programs. He identified three sources of potential difficulty when implementing change in a system.

The first source of potential difficulty is from people who do not believe in the suggested change. One PDS example would be attempting to change the teacher – professor hierarchy. Educational status quo would be having the professor tell the teacher what to do. To change this paradigm would be to upset the current accepted role of intellectual aristocracy in our society. The new PDS collaborative approach would be to have both the professor and the teacher learn from each other upsetting the status quo.

Resistance to a change in the status quo would be the initial response of people who do not believe in the suggest change.

A second source of difficulty is from people who do not want to change what they do. These people believe that it is important to improve student and teacher education but are more invested in doing what they have always done as opposed to making a change for improvement. Most classroom teachers prefer traditional practice over the uncertainty of educational innovation as required by the PDS[2] movement (Trubowitz & Longo, 1997). Currently, most university systems do not reward, assess, or promote faculty for consulting with schools and districts for instructional improvement which discourages the practice. Many university faculty members, especially junior faculty, find it necessary to remain on campus and commit their time and energy to their professional research agenda and the publication of articles in an effort to achieve tenure.

A third and final source of difficulty comes from people who believe that the PDS movement is important to the reformation of education but get frustrated with the slowness of change and the lack of time available to complete the work. Snyder (1999) also suggested that one of the biggest challenges of the PDS movement is to convince the public of its importance. The challenge is to create and maintain conditions that provide opportunities to transform skepticism into belief. Change occurs when institutions admit there is a problem and then explore and make changes to solve the problem and when the participants believe that they can make a positive difference for the institution through solving the problem (Snyder). One important way to help people and institutions to meet the three conditions for change is through continuous research. The research results can be used to help persuade people that a new approach or teaching method would be beneficial to them and their students (Snyder, Morrison, & Smith, 1996).

According to Darling-Hammond (2005), current PDS reform efforts within the schools and universities require exceptionally skilled teachers. Teachers entering the workforce must now possess the skills to motivate and educate all students to their highest level of academic performance. They must accommodate, celebrate, and respect student diversity while ensuring that all students learn to create, present, and synthesize their own ideas. To accomplish this end, teachers are required to have a clear understanding of learning as well as teaching while connecting student experiences with curriculum goals. More complex forms of teaching are required to support the wide range of learning styles and multiple intelligences encountered in the classroom. The ability to teach higher order thinking skills and to enhance student performance abilities are now expected from beginning and experienced teachers (Darling-Hammond, 2005).

Necessary outcomes needed from teacher preparation require new investment in teacher learning. Current PDS goals and objectives advocate for teacher autonomy while establishing connections and partnerships with parents and communities. To accomplish this, teachers must be able to predict and control consequences, connect theory to practice, and reflect on their work while being able to articulate and communicate complex ideas to both students and parents (Fenstermacher, 1992). Fraiser (1992) expressed the need for novice teachers' experiences to be based on Paulo Freire's (1970) notion of dialogue. The dialogue of teaching should take place in a respectful open forum where both novice and experienced teachers are encouraged to give voice to their teaching experience and reflections. New teacher educators with the required teaching skills will be the result of rigorous, sustained study while they maintain an on-going dialogue with master teacher educators (Fenstermacher, 1992).

Reform leading to growth of the PDS movement can be directly linked to the alternative certification movement (Dixon & Ishler, 1992). The PDS movement sought to strengthen school-university relations in an effort to provide an improved student teacher experience hoping to restore public confidence. The alternative certification movement was established in direct response to the teacher shortage in an attempt to ease the entry into the teaching field from other professions. The shortage of teachers coupled with the lack of public confidence led 41 states to provide some type of alternative certification route excluding the student teacher experience (Frazier, 1994). The argument used by legislatures to shorten or remove traditional student teaching is that on-the-job supervision is more effective in preparing teachers to teach. This truncation of the program saves training dollars and gets student teachers into the classroom more quickly. Wright, McKibbon, and Walton (1987) found that student teachers educated in a traditional teacher education program coupled with state-funded mentors outscored any of the recruits coming into the program through the alternative certification route. This research also suggested that there is an increase in teacher effectiveness and certification scores with the addition of mentoring for students in a traditional preservice program. Those preservice teachers participating in the alternative certification process appeared to lack needed classroom skills. Darling-Hammond (1992) found that teacher education programs with little or no student teaching leave recruits significantly underprepared in areas important to classroom effectiveness. In a proactive attempt by teacher education programs to avoid reacting defensively to increased regulation from legislatures, the PDS movement with lengthened student teaching opportunities and intense supervision was developed (Williams, 1993).

Another proactive attempt to improve teacher education later included in the PDS NCATE Standards was the idea of simultaneous educational renewal. It was first introduced in 1994 by John Goodlad who established the National Network for Education Renewal. Goodlad's network had two major foci: renewing the initial teacher preparation programs and simultaneous school renewal. Goodlad reported that schools that train student teachers and that are in the renewal process are analogous to teaching hospitals necessary for good medical education. For schools to be exemplary, teachers who teach in them must also be exemplary. He further wrote that if teacher education is to be exemplary, then the schools in which beginning teachers train must also be exemplary. In 1999, Snyder reported that a PDS is a district-school-university partnership dedicated to the improvement of K-12 education through simultaneous renewal of the education of classroom teachers and the institutions that train them. The idea of schools and university teacher education programs simultaneously renewing themselves through partnerships and professional development was folded into the PDS movement and considered an important strand (Teitel, 1999).

Currently, an authentic PDS must work diligently against existing traditional school norms in an effort to articulate and communicate a different set of expectations for student teachers and supervising teachers both in and out of the classroom. An authentic PDS goes beyond relationships between supervising teachers and student teachers and affects the overall professional practice environment. It encourages continuous inquiry, collaboration, collective work and professional collegiality. One important feature of an authentic PDS is the shared decision-making in teams within school and between schools and universities in an effort to discourage teacher isolation and to increase knowledge sharing, team planning, and collective reflection between and among all participating participants (Darling-Hammond, 2005).

One of the most likely outcomes of a strong PDS internship would be the reduction of the number of teachers leaving the profession during the first three years of teaching. Attrition rates for new teachers are very high; with most researchers finding that between 25-50% of them leave within the first few years of entering the profession (Guarino, Santibanez & Daley, 2006; Georgia Partnership for Teacher Education, 2006; Grissmer & Kirby, 1987). Many leave the profession because they have not felt they connected with their students, have experienced a sense of frustration or considered themselves failures in the classroom (Johnson & Bickeland, 2003). Currently, national teacher attrition costs over $2 billion annually (Fulton, Yoon, Lee, 2005).

Teacher efficacy is one of the strongest predictors of commitment to the teaching profession. It makes sense that providing additional support, such as teacher mentors coupled with a strong teacher preparation program, would help beginning teachers become more successful (Darling-Hammond, 2005). Beginning teachers often enter the teaching profession because of the need to make a difference in the lives of students. When these needs are met through quality training and exemplary supervisory teachers, everyone wins—students, teachers and the teaching profession.

Abdal-Haqq's (1998) review of the PDS internship literature concludes that novice teachers from PDS placements are better able to use varied pedagogical methods and practices in the classroom (Miller & O'Shea, 1994; Zeichner, 1992; Hallinan &

Khmelkov, 2001). The novice teachers are more reflective in their practice and are more knowledgeable of school routine and activities (Hayes & Wetherill, 1996; Trachtman, 1996). PDS trained novice teachers are more confident in their knowledge and skills as a professional and are better equipped to instruct ethnically and linguistically diverse student populations (Book, 1996). They are also more likely to take a full-time teaching job in an inner city school when they complete their practicum in urban areas and remain in the profession (Arends & Winitzky, 1996; Hayes & Wetherill).

Efforts have been made to standardize the definition and goals of a PDS. NCATE initiated a project in 1995 to develop standards for PDSs. Initially, the project was controversial with many objecting to the possible loss of creativity within the PDS movement if standards were identified. Ultimately, it was decided that the PDS movement needed rigor and accountability. Without articulated standards, the PDS movement might disappear. Obtaining a clear definition of what makes up a PDS was challenging because there was little agreement. Some PDSs were PDSs in name only, continuing to prepare teachers in the traditional manner. Other PDSs worked to improve teacher and student learning through a major shift in the student teaching process and in the collaborative relationship between the school and university (Levins & Churins, 1999). The NCATE (1997) working definition became

> collaboration between schools, colleges or departments of education, P-12 schools, school districts, and union/professional associations. The partnering institutions share responsibility for (1) the clinical preparation of new teachers; (2) the continuing development of school and university faculty; (3) the support of children's learning; and (4) the support of research directed at the improvement of teaching and learning. (p. 4)

This definition was developed from bottom-up wisdom and experience. The information was captured through an in-depth survey and a national conference. All advisory and working groups for the project had broad representation, including researchers, policymakers, teachers, and leaders of the PDS movement. From the national conference, the PDS standards emerged (Levin & Churins).

The NCATE developmental guidelines identify four different levels of developmental stages of a PDS based on five standards. Both the developmental guidelines and standards are clearly articulated in the 2001 NCATE document. These stages include beginning level, developing level, at-standard level, and leading level. During the beginning level, individual relationships are established and cultivated between the district-school and university. Mutual trust is beginning to develop and values are explored and discussed. It is during this phase that an overarching PDS philosophy begins to take shape. Early collaborative efforts are documented and verbal commitment from partners is obtained.

During the developing level, a formal agreement, usually a memorandum of understanding, is executed between the participants, and there is consensus on the mission and philosophy of the PDS. Partners are discussing the mission of the PDS partnership, and institutions are exploring changes in polices and procedures that would provide evidence of PDS institutionalization.

When a PDS reaches the at-standard level, the mission of the PDS partnership is integrated into the partnering institutions. Best practices are supported and used throughout the PDS seeking positive outcomes for all learners. Institutional policy and procedures have been changed and reflect the results of PDS lessons learned.

Sustaining and generative PDS work is found at the leading level. There are systematic changes in policy and procedures within the institutions that support PDS participants in meaningful ways. District, state and national impact is made based on researched PDS best practices from within the partnership. The differences between the developmental levels vary in the commitment level, PDS expertise, level of institutionalization and perhaps most importantly the impact the PDS partnership has on outside institutions (NCATE, 2001).

The five standards identified by NCATE (2001) are to be viewed holistically and often overlap. The characteristics of PDSs that are addressed by the five standards are learning community, accountability and quality assurance, collaboration, equity and diversity, and structures, resources and roles. These standards are used for both self-study and assessment team visits. Because PDSs are developmental in nature, a PDS partnership may develop unevenly among the standards.

The first standard, learning community, is very important to the development of a PDS. Development of P-12 students, candidates, and PDS partners through inquiry-based practice cannot be achieved without the umbrella of support provided by a committed learning community. Results obtained from inquiry-based practice leads to change and improvement in teaching pedagogy as well as the policies and procedures of the partnering institutions.

The second standard, accountability and quality assurance, focuses on upholding professional standards for teaching and learning both to the PDS partnership and to the public at large. Criteria for participation in the PDS are clearly articulated at both the institutional and individual level. Assessments are collaboratively developed and data collected in order to examine systematically the learning outcomes for P-12 students, candidates and faculty. There is impact through work in the PDS schools on policies and procedures at local, state and national levels.

The third standard, collaboration, helps PDS partners move from an independent, isolationist existence to an interdependent, cross institutional relational existence. Cross institutional partnerships encourage dialogue between K-12 and university faculty, which can bring about subsequent teaching and learning changes for both institutions. The potentially rich dialogue can bring theory and practice together for the improvement of teaching and learning across P-12 students, candidates and faculty.

Diversity and equity in learning is the fourth standard. PDS partners continuously review all policies and procedures to ensure equitable learning outcomes for the PDS community. Diversity is sought and explored with the PDS partners and learning communities. Equity, an even larger issue, is where the PDS movement seeks to obtain quality education for all students in the United States. The focus of the movement is to train highly qualified teachers to ensure this level of teacher competence in every classroom, along with adequate textbooks and supplies.

The fifth and final standard is structures, resources and roles. This standard encourages the establishment of governance and support structures necessary to undergird the partnership's mission. This would include systematic reviews of roles and responsibilities and the modification of them in order to achieve the PDS mission. Partners are also expected to broker resources from internal and external sources to support the PDS work. Communication among the PDS partners is crucial for the success of the partnership. It is imperative that PDS partnerships communicate effectively with all participants particularly the public, policy makers, and professional audiences of the work being done (NCATE, 2001).

If PDS schools are to continue to grow in number and gain credibility, they must also be visible and central to the plans of schools of education and school districts. Fundamental, continued support will be needed from national, state, and local levels. Implementation of PDSs is an expensive yet innovative way to approach teacher education. If PDSs are viable change agents, then they must prove themselves accountable to the public and parents of students alike (Sykes, 1997).

Research on Professional Development Schools

As the number of PDSs in the United States has grown, so has the number of studies being conducted. These studies address the theory, implementation, and description of PDSs, and many explore the nature and impact of district-school-university partnerships (Abdal-Haqq, 1988; Book, 1996; Compoy, 2000). Stallings and Kowalski (1990) found that while the number of PDSs have grown, there has been little systematic evaluation. In 1996, Abdal-Haqq reported that a majority of the PDS research being conducted has been focused on outcomes associated with preservice and inservice teachers. The research usually explores their satisfaction with teaching as a profession, teacher efficacy, perceived competence, and attrition. Book pointed out that most current PDS research is descriptive with little methodological detail included. The lack of detail raises questions about the transferability of the studies. With the lack of methodological detail, there is cause for concern over the ability to replicate the study. Abdal-Haqq identified a need for more research on the effect of PDSs as it relates to student academic achievement.

Because a PDS is involved in a variety of partnership activities, such as preservice and inservice teacher training, professional development of novice and experienced teachers, action research, and academic achievement of students, evaluation becomes difficult and complex. The difficulty in isolating PDS effects from other confounding variables makes it hard to determine if programs are clearly successes or failures (Compoy, 2000). Book's (1996) earlier research supports this premise. She stated that because of the complexity of the interactions that occur within a PDS setting, it is difficult for a researcher to account accurately for the impact of those interactions on the outcome variables being studied.

It is because of the difficulty and complexity of interactions that evaluators of PDSs tend to use qualitative methods that are primarily descriptive in nature. The methods used include interviews, questionnaires, surveys, journal writing, field notes, and classroom observations. In an attempt to understand the many educational connections, qualitative researchers have attempted to address the how and what questions connected to the study (Clark, 1999). There have been many case studies conducted in PDS schools. The case studies often focus on particular networks, such as National Network for Education Renewal (Osguthorpe et al., 1995) and the Benedum Collabrative (Hoffman, Reed, & Rosenbluth, 1997) or large collections of PDS studies (Darling-Hammond, 1994; Petrie, 1995). Thus, there are many case studies available for research review.

Book (1996) suggested the complexity of the school and classroom environments and the multitude of possible interaction factors cause descriptive methodologies to be predominantly used in PDS research. A thick description by the researcher of the complexity of interactions in the school and classroom gives the reader a clearer understanding of the PDS educational process. The use of descriptive methodologies allows the researcher to document the nuances in the evolution of partnerships between district, school, and universities. Ultimately, the PDS goal of promoting inquiry within the school setting is more conducive to the use of qualitative methodologies. However, there are a few quantitative examples of research in PDSs.

Teitel (2000) discussed the critical importance of systematic research to the growth and continued development of the PDS movement. Quality research, carefully constructed and implemented, allows stakeholders to assess the effectiveness and worthiness of implementation and maintenance of the PDS. Teitel stated that clear documentation of PDS effects is difficult because of a number of factors:

1. There is no universal agreement on the definition of "PDS." Looking for impacts on educational outcomes within settings that are not operationally PDSs might cause the evaluation to be flawed.
2. Collaborative partnerships are fragile. They could be damaged by a premature evaluation particularly because of the long-term nature of systemic changes.
3. Quantitative research using control groups is difficult to implement. Outcomes could be confounded by self-selection or program selection.
4. Stakeholders' perceptions can vary depending upon what outcomes are important and how to measure them.

5. Often outside evaluators are brought in because participants are actively working within the PDS. This can lead to a mismatch between the goals of the program and the direction of the evaluation.

Given these difficulties, it is not surprising that there are few credible and systematic assessments available to PDS researchers and stakeholders. Many of the studies that have been reported focus primarily on roles, relationships, creation of partnerships, teacher attitudes, and education in the PDS context. Logically, the next important research to be systematically conducted would be on the impact of PDS restructuring on student achievement. The identification of reform efforts and how they are related to the operationalization of a professional development school continue to elude researchers (Book, 1996; Reed, Kochan, Ross, & Kunkel, 2001). The inability to clearly define PDS treatments within a school is frustrating to the stakeholders and researchers. The complexity of the interactions of teaching and learning brings forward many issues that deserve in-depth research. The challenge lies in the difficulty of explaining and studying the impact of the complexity of the interactions, of teasing out the strands that directly affect student achievement and systematically researching them.

Educators are currently pressed by the business and political sectors to provide hard-data analyses of educational programs. Bottom-line results of sales and profits are a part of the process business people and politicians use in their respective sector practices. This same population insists that educators provide the same type of data concerning student pass rate, standardized achievement test scores, mathematics scores, communication scores, and retention rates for students. The business and political sectors also call for the outcomes of teacher education programs, including the number of teachers certified to teach, how long they stay in the profession, standardized test performance, performance as first-year teachers, and the impact teacher education graduates have had on the students they teach (Houston, Hollis, Clay, Ligons, & Roff, 1999).

Traditionally, educators have been reluctant to provide these types of data because of the complexity of factors involved. A variety of factors can affect student achievement in P-12, such as ethnicity, home environment, resources, and class size. Most educators cringe at the thought of overlaying a business model on a social system that must deal with diverse ethnicity, second-language learners, and multiple intelligences. The business approach to educational outcome analysis examines the products of teacher education programs as opposed to the process. For the past 15 years, researchers have attempted to link classroom teacher performance with student achievement (Gliessman, Pugh, Dowden, & Hutchins, 1988). In 1976, Rosenshine identified relationships between process variables such as specific teacher behaviors and product variables such as student achievement. Those behaviors include clarity, enthusiasm, task orientation, variability, and opportunities for students to learn concepts that are on achievement tests. In 1977, Medley reported that teachers who produce maximum achievement gains are more likely to enhance student self-concepts, and Powell (1978) concluded that teaching behaviors that are effective depend upon what content is to be learned. Pedagogical approaches differ from subject to subject.

More recent research conducted in Texas schools (Rivkin, Hanushek, & Kain, 2005) permitted the identification of teacher quality based on student performance. The results of this study revealed large differences among teachers in their impacts on achievement. It also shows that high quality instruction throughout elementary school can begin to compensate for a low socioeconomic background. Further results of the study support the notion that beginning teacher (years 1 – 3) perform significantly worse than more experienced teachers and that new teachers go through an adjustment period where the art of teaching is learned. It is at this point that 18% of Texas beginning teachers in this research study discovered that teaching is not the occupation of choice and subsequently leave the field.

Abdul-Haqq (1998) reported that there is little conclusive evidence that PDS programs improve student achievement. One possible reason for this is that inquiry and student achievement have been the two areas least systematically researched. The studies that have been conducted do not give a clear and concise description of teaching and learning activities that take place within the PDS program. Thus, the linkage between teacher development and student achievement has not been clearly identified and researched. According to Abdul-Haqq, the overall lack of convincing data is disturbing. It might spell the difference between continued growth of the PDS movement and its demise.

In response to the lack of achievement data, Teitel (2001) provided a current review of PDS research focusing on outcomes for preservice teachers, effects of professional development on experienced teachers, and the impact of the PDS model on student achievement. In an effort to respond to Abdul-Haqq's (1998) criticism of PDS researchers for the lack of data on student achievement, Teitel (2001) suggests that the body of research surrounding student achievement is growing. For instance, in 1995 Judge, Carriedo, and Johnson reported an increase in math score gains in one urban elementary PDS in Michigan from 3% satisfactory in 1991-92 to 48.3% four years later. Wiseman and Cooner (1996) describe dramatic increases in scores on the writing portion of a state achievement test through a PDS “writing buddies” program. Beginning with a 69% pass rate and increasing to a 92% pass rate, the principal of the school directly attributed the PDS partnership for helping increase the achievement of his students. Teitel pointed out the primary weakness of the study was the lack of an effective comparison group. Knight, Wiseman & Cooner reported an increase in the percentages of students mastering a state criterion-referenced test in mathematics for third and fourth grades. Prior to PDS implementation, mastery of the mathematics portion of the test were 70% and 64%. After implementation of the PDS intervention implemented by preservice teachers, the mathematics score percentage increased to 77% and 79%, respectively. Pine (2003) reported that an evaluation of a Michigan PDS can attribute an increase from 25.6% pass rate to a 97.8% pass rate over three years in mathematics scores on the state achievement test as a relentless emphasis on achievement in all aspects of the PDS.

In 2001, Teitel identified two of the most comprehensive and convincing large scale studies of the impact of PDSs. These two studies are The Benedum Collaborative Model of Teacher Education: A Preliminary Evaluation (2000) and The Houston Consortium (1992). This study is including Cooper and Corgin’s (2003) research examining Student Achievement in Maryland’s Professional Development Schools because it is a statewide initiative.

The Benedum Collaborative Model

The Benedum Collaborative Model study led by J. Webb-Dempsey of West Virginia University (WVU) primarily examines the effects of a PDS program implemented in 12 urban school systems on student achievement. Establishment of PDSs in the local public schools was coupled with a review of WVU's teacher education program during which the entire curriculum was re-invented. The following changes were made: (a) a move from a four-year bachelor's degree program to a five-year program that includes graduate school, (b) an increase in admissions requirements, and (c) assignment of each novice teacher to one PDS for three consecutive years.

There were three foci of the report. The first assessed quantitative data from WVU novice teachers and student records. The second involved student achievement at the PDSs, based on individual student test scores on the Stanford Achievement Test (9th edition), supplemented with additional school data published by the West Virginia Department of Education. Finally, interviews and site visits were conducted to consider the extent to which the PDSs are realizing less quantifiable goals of the collaborative. In addition, high school student and teacher surveys were conducted. However, no baseline data were collected, so the survey use is limited.

The Benedum Collaborative Model of Teacher Education was implemented at the College of Human Resources and Education (HR&E) at West Virginia University. WVU was one of the first universities to adopt the PDS model. HR&E worked to establish PDS relationships with 21 public schools, at the elementary, middle, and high school levels. The schools are situated within a five county area around Morgantown, West Virginia. There were three entry points in the PDS process requiring schools interested in becoming a PDS to apply and show faculty consensus on becoming a PDS.

The data include interviews with 400 students, surveys of 3,000 students; evaluation data collected by the State Department of Education and standardized achievement tests. The Stanford Achievement Tests (9th edition, SAT) scores for the previous three years were identified as the best available data. However, most PDSs participating in the study lacked baseline Stanford test score data from pre-PDS years. The 21 PDSs were compared with state and county averages of analyese of attendance, graduation rate and achievement test scores (Teitel, 2001b).

Webb-Dempsey (1997) reported that data analysis of the Benedum Collaborative Model of Teacher Education shared with schools and teachers often spurred them into action. Identified areas of concern such as playground behaviors and vocational technical track student attitudes encouraged faculty to collaborate and strategize ways to improve conditions for students within the classroom. Additionally, data provided information on the perception of how students learn best. Hands-on learning coupled with one-on-one instruction emerged as the method most students perceive as the most effective.

In 1999, mean basic skills scores on the SAT were higher in PDSs than in non-PDSs in grade six. The difference was statistically significant (.05 level) in grades six, seven, and eight. Additionally, scores on individual tests within the SAT show PDS students with increased gains with the largest gains in mathematics, although differences that have a substantial magnitude may not achieve statistical significance. At this time, the researchers felt it was more appropriate to examine the magnitude of the differences than the statistical significance (Gill & Hove, 2000). Thus, the discussion of gain difference focused on effect sizes rather than significant differences. The final analysis of data showed a slight increase in attendance and graduation rates with no significant difference in achievement (Teitel, 2001b).

The Houston Consortium

The most convincing large scale PDS study on the impact of PDSs is the Houston Consortium. The research used a quasi-experimental model combining test score data with qualitative observational methods. Using data obtained from four universities and three school districts, the study compared test scores of PDS and non-PDS students on the Texas test for certifying new teachers and student test scores on the Texas Assessment of Academic Skills (TASS). Classroom observations were also used to document instructional time on task. Consortium teachers were found to have higher certification test scores, spend more time checking student work and responding to students, encouraging student self-management, positive behavior and improving student performance (Teitel, 2001b).

The Houston Consortium study by Houston, Hollis, Clay, Ligons, and Roff (1999) was designed around six major objectives. These objects were (a) to create a consortium of diverse institutions in order to demonstrate the efficacy of shared governance and collaborative program development, (b) to design and implement a teacher preparation program based primarily in urban professional PDSs, (c) to provide professional development experiences for PDS and university faculties in response to their identified needs, (d) to integrate technological use for communication, management, and instruction, (e) to increase knowledge and performance of preservice teachers, and (f) to increase student learning. The final two objectives focused on outcomes of the programs.

The Houston Consortium was comprised of four urban universities, three school districts, and two intermediate school agencies. The study included the implementation of a redesigned urban teacher education program. The overarching program goal was to produce teachers who have the skill sets to be effective in culturally diverse and economically challenged environments of large urban school systems. Houston is the nation's fourth largest city and educates 20% of Texas school age population. This system also educates 30% of all African Americans, 40% of all Asian American and 16% of all Hispanic Americans in the state. Located in 16 elementary PDS schools, programs were implemented to achieve each of the six objectives in the demonstration program. The 16 were chosen because of the large number of students at risk of failure. These schools also mirrored the urban population of Houston and had faculties that were committed to improving academic achievement. Approximately 14,000 students attended these schools with 33 student teachers participating (Houston, Clay, Hollis, Ligons, Roff & Lopez, 1995).

Although there were six objectives, the focus of the sixth objective was to improve standardized test scores of P-12 students. This objective was assessed through analysis of changes in achievement of PDS students on the Texas Assessment of Academic Skills. TAAS scores were analyzed in mathematics, reading and writing at the fourth, eighth and tenth grade levels from 1992-93 through 1994-95. The 1992-93 TASS were administered before schools became PDSs and can be considered as baseline data.

Of the 16 participating PDS elementary schools, 14 showed an increase in reading with 2 showing a decrease. In mathematics, all 16 PDS sites showed an increase. As measured by the TAAS, writing skills increased in 10 of the schools and decreased in 8 of the schools. In these schools, preservice teachers taught math and reading to small groups and individual students, but they did not teach writing.

The Houston Consortium research concluded that there were significant positive changes in student achievement over the two-year period. It appears that during the first year of a school's becoming a PDS, the changes were the greatest. During the second year, achievement gains appeared to stabilize, but were still higher than scores in the school before becoming a PDS. Several factors were identified as having some effect on achievement gains:

1. Because of the placement of pre-interns and interns, there was a large number of adults available to instruct students.
2. Professional development was made available to teachers in the areas of urban teaching strategies and the use of technology in the classroom.
3. There was a school/university leadership network available to help students and teachers solve problems as they surfaced.
4. There was a possibility that the Hawthorne Effect was a factor in influencing student achievement.

The Houston Consortium conducted both formative and summative assessments on the program's impact on student teachers and K-12 student achievement. The overall research results tended to be supportive of program effectiveness in both educating teachers and student achievement (Houston et al., 1995)

Maryland's Professional Development Schools

This study by Cooper and Corbin (2003) was purposely designed to measure the effect of Maryland's state-funded professional development schools program implementation on student achievement. It was also designed to help fill the gap in empirical knowledge of PDS effects on student achievement because there are few studies currently addressing this issue. Also, this study's result was constructed such that a conservative estimate of the PDS effect could be obtained within a controlled study. The Maryland School Performance Assessment Program (MSPAP), the state-mandated achievement test, was used to measure student outcomes.

In 1998, 12 schools were chosen from 30 listed in the Maryland State Department of Education (MSDE) directory of PDSs. To qualify for state funding, each PDS was required to submit a detailed partnership plan, which was then reviewed by the state education department. Annual progress reports were submitted to the state ensuring that elements of the PDS were being met as required by MSDE. These elements included (a) school and campus-based preservice teacher preparation, (b) continuing education for school and university-based faculty, (c) integration of preservice and inservice components, (d) inquiry-based strategies for continuous assessment and improvement, and (e) substantial emphasis on teaching and learning in diverse and disadvantaged schools. An additional criterion used to identify the 12 participating PDS schools was that the PDS be completing at least its 4th full year of implementation by May 2000.

Once the 12 participating PDS schools were identified, a control group was formed. The goal was to achieve a comparison sample of non-PDS schools, matched on demographic data to match the 12 PDS schools including eligibility for free and reduced lunch, ESOL participation, and mobility of students. The demographic comparability was of the control school match was verified through knowledgeable school personnel in each participating PDS.

Student achievement data were obtained from the state's website, with the scores being reported as percentages of students in each school who achieved Satisfactory or

Excellent levels of performance on the six MSPAP subtests: reading, writing, language usage, mathematics, science, and social studies. Because of the way state data are reported, the unit of analysis is the school (grades 3, 5, and 8). Only satisfactory level data on 3rd grade students were analyzed. Fifth and eighthth grade students would have only been partially exposed to the PDS experience; thus, these data were not analyzed. The third grade data were used because in most cases the students would have been exposed to the PDS treatment during the majority of their school years, making the contrast between PDS and non-PDS schools easier to interpret.

To identify PDS effects on achievement, the statistical tests of differences between PDS schools and non-PDS schools MSPAP performance were designed to detect trends over a seven-year period. Alpha was set at .05. The trends were then compared for linear as well as nonlinear effects over the seven-year timeframe. It was decided that a year-by-year comparison would be less stable than a seven-year timeframe because of the possibility of random fluctuations.

When compared to state results, the data were similar on MSPAP performance in both PDS and non-PDS sites. Across the six MSPAP subtests, state averages ranged from 39.2 in reading to 49.5 in writing. Comparable scores for PDS schools ranged from 38.3 in reading to 50.1 in writing. Non-PDS schools scored in a range from 44.7 in social studies to 52.7 in writing. No means were given.

When seven-year trends were examined, the results did not show superiority of either a PDS or non-PDS school. The primary tests of PDS effects over time were analyzed as multivariate comparisons between the PDS and non-PDS seven-year trend averages. Tests of year by type interactions did not show significant differences between groups.

In additional to increasing student achievement in the PDS model, the Maryland scope included the following university activities: (a) identifying standards and standards-based assessments in teacher preparation, (b) applying standards and assessment in professional development, (c) integrating new-teacher preparation and ongoing professional development to achieve congruence and to alleviate existing tensions among participants, (d) reforming curriculum in both schools and universities, and (e) encouraging action research and inquiry processes in a systematic effort to gain a clearer understanding of the teaching and learning process. Student achievement is only one component of the complex PDS mission. Until the entire program is evaluated, the results remain incomplete. Initially, the Maryland PDS design focused primarily on inservice and preservice teacher education. Only after student achievement was made the primary focus of PDS refunding, additional data then began to be collected (Cooper & Corbin, 2003).

Cooper and Corbin (2003) also report their concern about using only one method of assessment. They recommended the use of alternative assessment including student portfolios, journals, report cards, as well as teacher observations and evaluations and curriculum-based assessments tied to specific instructional designs. Because instruction in PDSs is more child-centered, the use of standardized testing as the measure of PDS outcomes is questioned (Ross, Brownell, Sindelar, & Vandiver, 1999). Webb-Dempsey reported in 1997 that traditional standardized testing frequently conflict or ignores child-centered practices. When standardized tests are used, often content and methods are tied to them. Classroom teachers have little latitude in changing either (Abdal-Haqq, 1998). Student teachers in both PDS and non-PDS sites express frustration when they are restrained from using culturally responsive practices, strategies, or content because of the system-required focus on standardized test preparation (Wiseman & Cooner, 1996). Many advocate the use of additional methods of assessment however; researchers have been slow to develop alternatives (Abdal-Haqq, 1998).

Need for More Student Achievement Research

Student achievement is rarely the focus of PDS research. Teitel (2001) stated that impact studies on academic achievement in PDSs are beginning to appear. One annotated bibliography and several articles based on review of publications have been published in an effort to determine the areas where the preponderance of research has taken place. Abdal-Haqq's (1993) annotated bibliography of primary PDS resources included 119 listings. Of those listings, only 33 were identified as research reports. The primary resources review by Abdal-Haqq included both published and unpublished literature. Papers written for presentation, handbooks, bibliographies, course outlines, policy statements, and historical perspectives were included. In 1993, Valli, Cooper, and Frankes, using the Abdal-Haqq annotated bibliography as the base, conducted their own review of PDS literature. Using only original research, Valli et al. found that of the 59 studies reviewed, 20 focused on key participants in a variety of roles and sites, 14 focused on elementary schools with six focused on middle schools. Five of the studies were conducted in high schools, 5 in urban settings, 6 encompassed a review of the PDS model in general, and 3 were individual reports on rural settings, state departments, and Holmes Group members. While all of the educational methods being studied either directly or indirectly impacted student achievement in some way, the greatest focus of PDS research as documented by Valli et al. has been on collaborative relationships between university and K-12 systems (23 studies). This group also reviewed 14 studies reltated to professional development teacher education, the organization and structure of schools, the nature of teaching and learning, and the process of inquiry. When Valli et al. published in 1993, no study reviewed was identified as focused on student achievement. This study also reviewed data on the achievement of low socio-economic status students and students of color. The Valli et al. review showed little if any evidence of PDS impacts on student achievement at all.

In 1998, Teitel described his 1995 search of 200 PDS documents. The PDS documents included: 15 books and 19 items such as handbooks, 86 descriptions of PDS programs, 41 policy or opinion pieces, 18 surveys, and 18 case studies. In 2001, Teitel reported that systematic research on PDSs was virtually non-existent. In 2004, Teitel updated the NCATE professional development schools review of research. The section on student outcomes identified 14 exemplary studies. These studies attempted to make connections between changes taking place in the PDS and student achievement primarily through the use of standardized test data.

Pine (2003) reported single site longitudinal data for an elementary school Pontiac, Michigan. When test scores were compared with other schools in the district and state school scores, the Pontiac Elementary School, over an eight-year period, met or exceeded the state and district averages. Pine attributes their success to a focus on student learning by the PDS. Frey’s (2002) two-year study of an urban San Diego PDS school reported an overall increase in reading scores of 31% on the SAT-9 tests with 55% of seventh graders scoring above grade level. Eighth graders also showed a remarkable increase in reading portfolio scores over a two-year period going from 18% to 70% above grade level. This school was comprised of 48% ESOL learners, with all students qualifying for free or reduced lunch. Frey credited the school's success in student literacy on the development of learning communities. Through detailed documentation, Frey was able to connect learning experiences with academic gains.

Teitel's (2004) reported that a matched control research design was used by Castle, Arends, Ware, Rockwood, & Deniz to further inform PDS participants. In this project only school-wide aggregated standardized test scores were analyzed. Castle et al. reported difficulties in using standardized test scores as the only outcome measure. They also found that there were no significant differences between the PDSs and non-PDSs when aggregated data were used. The study found that socio-economic status and ethnicity were much stronger predictors of academic success than that of attending a PDS school. Teitel goes on to report that in another study by Castle and Rockwell (2002), the use of aggregate data can cause researchers to overlook some important impacts.

Teitel (2004) reported that a Kansas State University research study has made a clear connection between student academic achievement and the level of PDS attainment. In the study by Yahnke, Shroyer, Bietau, Hanclock, and Bennet (2003), student achievement test scores in mathematics were tracked over multiple years. The PDS scores were compared to state averages and to each other based on length of time as a PDS, self-assessment ratings aligned with the NCATE PDS standards, learning communities, and faculty engagement. This study showed a 19% gain in the oldest third of the PDSs as compared with a .7% gain in the newest PDSs. The middle group of PDSs showed a 26% gain. The researchers discovered that the level of faculty involvement was more important than the length of time a school had been a PDS. Schools reporting a high level of faculty engagement showed a 23% gain over schools with low faculty involvement at 3%. The report also included a description of student academic achievement improvement in a low-performing, high-poverty school that showed an annual gain of about 30% which was triple the state average.

As the PDS movement grows and matures, research is being conducted more scientifically in an effort to show student academic achievement impact or lack thereof. The movement toward a more scientific approach supports continuous efforts to improve the quality and persuasiveness of all educational research (Cook, 2002; Riehl, 2006; Slavin, 2004). Since enactment of the *No Child Left Behind Act* of 2001, there is now a stronger focus on student learning outcomes throughout the American educational system as well as in the PDS movement. The use of high stakes standardized tests has forced the American educational system to demonstrate measurable student outcomes when attempting to validate new innovative programs or models. This new accountability (Carnoy, Elmore, & Siskin, 2003; Fuhrman, 1999) places the impetus upon student performance with the locus of responsibility for performance directly on local schools. The growing expectation that local schools meet the academic needs of a much more diverse group of students to much higher predetermined standards creates much greater demands on teachers (Darling-Hammond, 2000). The new accountability also includes the use of public reporting to student outcomes, rewards, and sanctions to encourage change in failing schools curriculum and instructional practices (Fuhrman, 1999; McDonnell, 2004). NCLB goes beyond the accountability policies of many states by attaching significant stakes to individual school performance. Additional stakes are possible school restructuring and the threat of closure for schools that continue to fail to meet accountability targets. To avoid federal sanctions, schools must make adequate yearly progress toward predetermined proficiency benchmarks. NCLB assumes that external federal pressure will encourage districts and school to work together to identify ways to improve student achievement (Goldrick-Rab & Mazzeo, 2005). Thus, standardized student achievement tests have become an important part of educational assessment and are increasingly becoming seen as the critical outcome measure for student learning (Teitel, 2004).

There are several student characteristics or factors that have been identified as having a systematic impact on student achievement. These factors are socioeconomic status (SES), gender, ethnicity, and English as a secondary language. Grusky (1994) reported that stratifying societal forces have begun to shift away from class-based variables, such as SES, to ethnicity and gender. This study will use ethnicity as the variable to identify and show proportion of change in student achievement gaps.

In 2000, approximately 40% of the national enrollment for U.S. public schools came from different ethnic backgrounds. During the past 30 years, a significant number of ethnic minority students have scored lower on standardized tests when compared to their White peers (National Center for Educational Statistics, 2002). Several recent studies have found that ethnic factors can affect educational outcomes with students from the main ethnic minority groups achieving below the average for that of their White peers (Demie, Reid, & Butler, 1997). Griffin (2002) reported that low academic performance of minority and disadvantaged children continues to be a persistent problem in U.S.

education. Numerous researchers have attempted to explain the differences in academic achievement for children from different racial and ethnic backgrounds. Others have attempted to identify successful interventions that lessen differences in achievement in children from different racial and ethnic backgrounds (Griffith, 2002). It is important that implementation of the PDS model and subsequent research studies explore the effect the model has on achievement. This is particularly important research on children from different racial and ethnic backgrounds. Identifying achievement gaps and successful PDS interventions through the use of quantitative data analysis will help prove or disprove the success of PDS models. Exploring ethnicity and ESOL within this PDS model may show that a positive school environment coupled with an experiential curriculum may help compensate for the negative effects of low-SES (Scales, Roehlkepartain,Neal, Kielsmeier, & Benson, 2006).

Identifying achievement gaps between minority and non-minority groups has long been an issue for educational researchers, politicians and educators (Bainbridge & Lasley, 2002). Typically this meant that much of the literature focused on the differences between African American and White students because they have traditionally been the largest groups leaving other minority populations such as Hispanics or Latina/os under researched (Carpenter, Ramirez & Severn, 2006). This deficit in minority gap research makes it appear that many researchers assume that achievement gap causations are the same or similar for all minority groups (Bowman, 2001). It is clear that African Americans and Latino populations' achievement gap causations are different in the academic setting (Lee, 2003). Conceptualizing the achievement gap as only one minority population ignores between-group differences but more importantly ignores within-group differences (Gutierrez & Rogoff, 2003). Identifying the multiple achievement gaps within and among the five major ethnicities within a PDS setting is a goal of this research.

Closing the achievement gap will not be easy. From an educational standpoint, it will require new approaches to understanding demographics, diversity and accountability. It will require that all educational entities and agencies get a better understanding of the social complexity of the educational process (Bainbridge & Lasley, 2002). Schools should move toward a focus on the educational diversity of the student population and the intellectual capital that these students bring to the classroom. Finally, schools should seek to embrace rich, creative, academically sound educational programs designed to meet the needs of an ever-changing diverse student population.

CHAPTER 3

METHODOLOGY

The goal of this study is to explore the impact of the Professional Development School Deliver Success (PDS^2) model on student outcomes by comparing student achievement scores. Student achievement scores in mathematics and science from twelve PDS schools are compared to twelve demographically matched non-PDS schools from four metropolitan school districts. The PDS schools have been matched with comparison schools in the same system on proportion of students on free or reduced lunch, on the previous year's academic achievement and on the proportion of members from different racial and ethnic groups composition. The research questions being addressed are the following:

1. How does the PDS^2 model affect mean student achievement in mathematics and science as measured by the Georgia Criterion-Referenced Competency Tests and Georgia High School Graduation Tests?
2. Are there significant differences in mean achievement test scores between PDS^2 feeder pattern schools and control schools?
3. From the baseline year to the end of the first year, how many PDS_2 and comparison schools have changed their Adequate Yearly Progress status and in what direction?

4. Is there a mean difference between ethnic groups on the scaled scores of the CRCT and HSGT for mathematics and science?
5. From the baseline year CRCT and HSGT to the end of the first year for across ethnic groups, is there a mean difference in scaled scores on the CRCT and HSGT for mathematics or science?
6. From the baseline year CRCT and HSGT to the end of the first year for different ethnic groups, is there a correlation between the proportion passing change on the CRCT and HSGT for mathematics or science and Hedges's g effect size?

In this study untreated control (comparison) group with separate pretest and posttest samples were used to explore the effects of the PDS^2 intervention on student achievement. A quasi-experimental design was chosen because a true experimental design was not feasible for use in the school setting. A primary reason true experimental design is not feasible is because subjects are assigned randomly to comparison and treatment groups. This is a necessary condition for true experimental design that cannot be met in a school setting. Another reason for using a quasi-experimental design is that a true experimental design may be too costly, time consuming, or it may presume the ability to manipulate an intervention that has already occurred. The use of a quasi-experimental design is usually the best alternative for maximizing internal validity (Schutt, 1996). A quasi-experimental design is one where the control group is comparable to treatment group in predetermined ways. Two frequently used predetermined ways are being eligible for the same services or being in the same school group (Rossi & Freeman, 1989). The quasi-experimental design is the most frequently used in social science research because it meets the conditions necessary for conducting research in a school setting (Shadish, Cook, & Campbell, 2002). In designing the research, threats to the validity of the research design were reviewed (Shadish, Cook, & Campbell, 2002). A possible but unlikely threat is instrumentation; however, the State of Georgia testing personnel assured the researcher that the CRCT and HSGT scores are horizontally equated within tests developed for both the QCC or GPS curriculum.

Another threat given attention was historical. An example of a historical threat to internal validity would be the introduction of a new academic support model such as after-school tutoring or chess club that is available to a particular population of students in a PDS school. Students who participate in extra-curricular academic programs are subject to the influence of the program that is not a part of the PDS program. Thus, it would be difficult to identify and definitively say which intervention was directly linked to academic achievement. Because the researcher was invited by the system into the school, there can be no control of programs added by the school administration after the study has begun.

A third threat would be that of maturation. This threat arises when students in one group are growing more experienced, tired, or bored than students in another group. This can occur between and among PDS^2 and comparison school classrooms and often reflects on the quality of the teacher. However, given the number of schools and teachers, it is unlikely that this threat would be operating differentially between the set of PDS^2 and comparison schools.

Participants

Sample data for this study were collected during the baseline and year one of the PDS^2 project implementation. Data were collected from twenty-four schools in four of the largest urban metropolitan area school systems in the southeast. The data set included 26,529 students with 8,053 students in elementary school, 12,969 in middle school, and 5,507 in high school. The data were collected from 12 PDS^2 schools and 12 non-PDS^2 schools. Four feeder pattern school sets were randomly chosen to include one elementary, one middle, and one high school in each of the four systems. The sample size of well over 250 per group gives a power of .99 for a medium effect size of .5 for testing mean differences using an independent t-test with $\alpha = .05$ (Cohen, 1977).

Students in 24 schools, 12 PDS^2 and 12 comparison schools constructed in feeder patterns were measured on achievement and demographic variables. Mathematics and science test scores were the dependent variables with year, treatment, and ethnicity, as between factors or the independent variables. In each ANOVA, the dependent variable was a scaled score on the CRCT or HSGT. The science and mathematics scores were analyzed using an ANOVA rather than a MANOVA because not all grade levels from first to eighth were administered the science test. Thus, either there would be substantial missing data on the science dependent variable or using casewise deletion in SPSS several grade levels of mathematics scores would be deleted. A change in mean test scores within a given year for levels of the factor of ethnicity were analyzed as a preplanned test. Also, for a given ethnic group, year-to-year changes in mean test scores were analyzed.

Instruments

The CRCT and the HSGT are the two instruments used to measure how well students acquired the academic skills described in the QCC and the GPS. These assessments provide information on academic achievement at the student, class, school, system, and state levels. The academic information obtained is then used to gauge the quality of education throughout the state of Georgia. The academic information is also used to identify learning strengths and weaknesses of individual students as related to the QCC and the GPS. The CRCT is required for all students in grades one through eight. However, the science portion is only given in grades two through eight. The students take content area tests in reading, English/language arts, and mathematics. Criterion-referenced tests are specifically designed to measure how well students acquire knowledge and skills as stated in a specific curriculum or unit of instruction. The CRCT only assesses content standards outlined in the QCC and GPS (Georgia Department of Education, n.d.a).

An additional purpose of assessment is to provide information to parents, students, teachers, administrators, and legislators for use in educational decision making. In an effort to provide meaningful information, performance standards (cut scores) were established that systematically identify the level of proficiency of individual students and groups of students (Georgia Department of Education, n.d.c). Performance standards, identified by a preassigned cut score, passing score, pass-fail score or mastery score, essentially identify the number of questions a student needs to get right on a large-scale standardized test in order to pass. On the CRCT, all scores are reported as scale scores, which can range form 150 to 450 for each grade and content area. Each content and domain area has a scale score reported for it. Scale scores use a statistical process to convert the number of answers correct (raw score) to the CRCT scale using the Rasch model. Scores that are at or above 350 are used to indicate an academic performance that Exceeds the Standard set for the tests. Scores from 300 to 349 are used to indicate an academic performance that Meets the Standard set for the test. Scores below 300 indicate an academic performance that Does Not Meet the Standard set for the test ("2006 CRCT Interpretive Guide," Georgia Department of Education, n.d.a).

Passing the HSGT is required by all students seeking a Georgia high school diploma. The test content areas are English/language arts, mathematics, science, social science, and writing. These assessments are used to ensure that students qualifying for graduation have mastered QCC and GPS skills. Students who do not pass on the first administration of the test are given multiple opportunities to receive remedial instruction, retest and qualify to graduate before the spring of their 12th-grade year. Each content area has a pass and pass plus score assigned to it. A score of 500 on all four of the content areas receives a pass. Pass Plus scores are 538 for English/language arts, 535 for mathematics, 531 for science and 526 for social studies. (Georgia Department of Education, n.d.c.).

Procedures

Six ANOVAs, with three factors each, where ANOVAs at the elementary, middle, and high school levels were run separately for two different dependent variables. These were conducted to determine the effect of the PDS^2 model over the four participating school systems. There are two reasons the decision was made to report the data at the system level. The first reason was to keep the data at a high level of ambiguity in order to maintain confidentiality. The second reason was that differences among school systems were not at this point a major research question.

It should be noted that the PDS^2 and comparison schools from each system were matched and were high needs schools. Thus, it was believed that the matched pairs of schools were similar. However, it could be argued that just the process of matching implies that a blocked design would be more appropriate. The probable, although not necessarily, reduction in the error term from blocking would be most likely be compensated for due to the very large sample sizes. As almost always with the Null Hypothesis Significant Test, the reader should be cautious about interpreting rejection or failure to reject the null hypothesis when the p-value is near the alpha level.

The three factors used in the six ANOVASs are year, treatment, and ethnicity. If statistical significance for main effects or interactions is found, then a detailed breakdown of the means for different levels of a factor or combination of factors in an interaction is desirable in order to better understand which factors or combination of factors are responsible for differences in means. The follow-up test used in this research was the Tukey B test. This analysis was used because pairwise differences are often the most meaningful and easy to interpret (Stevens, 2002). The Tukey B uses the studentized range distribution to make pairwise comparisons between groups. The critical value is the average of the corresponding value for the Tukey's honestly significant difference test and the Student-Newman-Keuls test (SPSS Inc., 2005).

Additionally, I computed the effect size for all between group differences identified. This statistical procedure was used because the APA Task Force on Statistical Inference (Shea, 1996) reported that the educational research field as a whole has been moving away from using only statistical significance tests. The Task Force goes on to report that there is a movement placing emphasis on practical significance and the ability to reproduce the results. While the American Psychological Association style manual (APA, 1994) encourages that effect size information be provided, this has not led to changes in the way statistical reports are written (Kirk, 1996). Wilkinson and the APA Task Force on Statistical Inference (1999) reported that it is essential to good research that effect sizes be reported and interpreted in relationship to previously reported effects.

There are three important reasons for including effect sizes in research findings. The first important reason for providing effect size is that it facilitates inclusion of the article in subsequent meta-analyses. Second, it creates a basis from which researchers can include prior conclusions in subsequent articles published on the subject. The effect size information allows the researcher to formulate and design more specific research study expectations. Finally, inclusion of effect size facilitates the evaluation of how a study fits into existing literature. Effect sizes, when reported, highlights the similarities or differences in related research giving the researcher a place from which to judge practical significance (Kirk, 1996, Wilkinson & APA Task Force on Statistical Inferences, 1999, Thompson, 2006).

If research expectations match the null hypothesis when the null hypothesis specifies no difference, the effect size would be zero within sampling error. However, if the expectations do not match the null hypothesis, the expected effect size would not be zero. In general, effect sizes are an average statistic for the particular set of data being analyzed (Vacha-Haase & Thompson, 2004).

In 1976, Glass proposed an estimator of effect size based on the sample value of the standardized mean difference divided by the control group standard deviation. Hedges and Olkin (1985) argued that the assumption of equal population variances suggests that the population estimate is best obtained by pooling. Therefore, I used Hedges's g to estimate effect size:

$$g = \left[1 - \frac{3}{4N-9}\right] \bullet \left[\frac{\overline{X}_1 - \overline{X}_2}{S}\right],$$

where $\overline{X}_i$ is the mean of the *i*th group, S is the pooled standard deviation, and *N* is the total sample size.

> As N gets large, J(N-2) [a correction factor] approaches unity, so that the distributions of both [Hedges's g] and [Cohen's d] tend to a normal distribution with identical means and variances in large samples. Since [g] tends to [d] in probability as N gets large, [g] and [d] are essentially the same estimator in large samples. (Hedges & Olkin, p. 81)

The endpoints for the confidence intervals used for Hedges's g in this dissertation are based on the large sample properties of this estimator for an asympototic variance which are believed to produce reasonably accurate estimates (see Hedges' & Oklin, pp. 86-87). For small samples, when an effect size is present, a noncentral t-distribution is one of the preferred methods of calculating the endpoints for an effect size confidence interval.

Suggestions for interpreting effect sizes have been presented by Cohen (1968). He proposed that the terms “small,” “medium,” and “large” are relative to each other and to the research method being used in any given investigation. A common conventional frame of reference has been recommended by Cohen when no better basis for estimation of effect size is available. *Small effect size: d = .2.* Effect sizes are likely to be small because the research is being conducted in uncontrolled setting with uncontrollable variables. *Medium effect size: d = .5.* A medium effect size would be one that through normal experience an average difference would be noticeable to the researcher. *Large effect size: = .8.* This would require that the two populations be so separated that almost half of their areas are not overlapping. Another example of this distribution would be where 65.5% of the highest population exceeds 65.5% of the lowest population (Cohen, 1987).

The context of a study strongly affects the evaluation of the effect size. Vacha-Haase and Thompson (2004) reported that effect sizes should be interpreted by first considering what is being studied and by reviewing effect sizes of similar studies. This review allows evaluation of the replicability of the study. Cohen (1994) pointed out that statistical significance tests do not evaluate results replicability. Finally, comparing effect sizes across studies allows unusual results to be more easily identified (Vacha-Haase & Thompson).

Identifying unusual effect sizes essentially comes down to a value judgment of the researcher based on context, type of research, and practical significance. Such value judgments are and have always been a part of the researcher's job. In 2000, Tracey suggested that including effect size makes the research results easier for most people to grasp than p values. Inclusion of effect sizes along with confidence intervals provides added value to the research results and allow for easy inclusion in meta-analytic studies.

Because of the increased research value of reporting both effect sizes and confidence intervals, researchers have recommended that all results, including those that are statistically nonsignificant, be included in research reports. Not reporting effect sizes for nonsignificant results means treating these results as zero, potentially leading to falsely accepting the null hypothesis (Vacha-Haase & Thompson, 2004). All effect sizes and confidence intervals should be provided when p-values are reported (Wilkinson & APA Taskforce, 1999). Effect sizes and confidence intervals have been slow to be included until recently when more statistical packages have begun to include effect size computations within them (Thompson, 2002).

Confidence intervals also help to interpret research results when compared to previous research in a similar field. A review of previous research would encourage the researcher to reflect on the causation of differences in their own research vs. prior research. It would, in effect, enable the researcher to judge the preciseness of the research design or lack there of (Vacha-Haase & Thompson, 2004).

The effect size and confidence interval for all significant and nonsignificant results will be calculated. Effect sizes and confidence intervals identified as significant are reported within the dissertation. Confidence intervals identified as non-significant are available in Appendix A.

CHAPTER 4

RESULTS

Data from 24 schools in four urban school systems were collected. The data included CRCT scores for all students in each of the 21 elementary and middle schools. It included HSGT scores for all 11th grade students in four high schools. Because of the size of the data set, it was divided into six subsets by content area (mathematics or science) and by school level (elementary, middle, or high). These divisions allowed for a more organized and systematic data analysis. The dependent variables were mathematics and science scaled scores. The independent variables were year, treatment, and student's ethnicity. All school data were analyzed in both mathematics and science. The data were further disaggregated by ethnicity in an effort to identify mean gains or losses in achievement across and between groups.

Research Question 1 addresses the effect of the PDS^2 model on mean student achievement for mathematics and science across 1 year (see Table 1). When analyzing data collected for PDS only, I found three of the six ANOVAs to be statistically significant: elementary schools with science as the dependent variable, middle school with mathematics as the dependent variable, high school with mathematics as the dependent variable. However, the overall results indicated no statically significant gains in mathematics and science means for the PDS^2 schools in relation to the comparison schools for the first year of PDS^2 implementation.

Table 1

Analysis of Growth Over Year

Professional Development Schools

	2005			2006						
Subj.	M	SD	N	M	SD	N	F	p	MSE	d
Elementary School Level										
Math	324.90	30.82	2279	325.13	29.03	2180	0.062	0.803	897.66	0.008
Sci.	314.51	23.55	1303	317.22	22.51	1204	8.656	0.003	531.48	0.118
Middle School Level										
Math	312.70	31.65	2732	316.37	29.13	1031	10.501	0.001	959.62	0.118
Sci.	311.85	22.34	2732	312.31	23.61	1031	0.306	0.580	515.12	0.020
High School Level										
Math	517.83	22.66	1143	522.19	26.71	1145	17.728	0.000	613.62	0.176
Sci.	497.12	24.96	1161	497.56	19.53	983	0.202	0.653	512.31	0.019

Comparison Schools

	2005			2006						
Subj.	M	SD	N	M	SD	N	F	p	MSE	d
Elementary School Level										
Math	329.12	32.88	1760	330.82	30.77	1814	2.538	0.111	1013.00	0.053
Sci.	315.18	26.09	1039	319.24	22.34	1050	14.608	0.000	589.32	0.167
Middle School Level										
Math	309.36	29.20	2505	312.53	29.31	953	8.084	0.004	854.37	0.108
Sci.	310.39	20.33	2505	308.19	21.19	953	7.907	0.005	423.21	-0.107
High School Level										
Math	522.75	34.16	1261	528.56	29.19	1445	22.733	0.000	998.84	0.184
Sci.	505.40	23.82	1253	496.65	19.49	919	83.083	0.000	488.00	-0.396

Note. F is F-test value for year effect in ANOVA. p is tail probability for F-test value. MSE = Mean Standard Error. d = effect size calculated by subtracting means and dividing by the square root of the MSE.

In the subsequent sections, the findings from the six ANOVA's which address research question 2 are discussed. At the end of the chapter, the six research questions are presented along with a summary of the results which are used to answer the questions.

Elementary School Mathematics

Elementary CRCT scores were collected from the four urban schools systems for all PDS and comparison school students participating in the research project. An ANOVA was run to identify significant differences between the first and second year, ethnicity, and PDS/comparison schools in the means. The significant differences identified by the F tests in the ANOVA with CRCT mathematics as the dependent variable were treatment, ethnicity, and treatment by ethnicity (see Table 2). The statistical significance of ethnicity by treatment is presented in Figure 1. The Tukey B test showed no significant difference in the means between African American students' and Hispanic students' achievement (see Table 3). While there was a difference between the means for African American students and the means for Hispanic students, results from Asian/Pacific Islander students, multiracial students, and White students showed no significant difference across all means.

Elementary School Science

An ANOVA was run to identify statistical differences between the means for PDS and comparison school students in the four systems that participated. The statistical significant differences identified by the F tests in the ANOVA with CRCT science as the dependent variable were student ethnicity and treatment (see Table 4). The statistical significance of ethnicity by treatment interaction is shown in Figure 2. There are multiple possible explanations for the interaction in Figure 2 that occurs in all six ANOVAs. One possible explanation is that the matching on ethnicity was made based on 2004 CRCT data. At that time it was not an effect answering a question asked in the research, but it could be conceptualized as a source of variation in the ANOVA tables. In reviewing the

Table 2

Tests of Between-Subjects Effects for Elementary School Students' Mathematics Scores

Source	Type III Sum of Squares	df	Mean Square	F	p
Corrected Model	359233.658*	19	18907.035	20.720	.000
Intercept	221267181.733	1	221267181.733	242487.848	.000
Year	19.621	1	19.621	0.022	.883
Treatment	14847.621	1	14847.621	16.272	.000
Ethnicity	237168.809	4	59292.202	64.979	.000
Year * Treatment	121.753	1	121.753	0.133	.715
Year * Ethnicity	1768.478	4	442.119	0.485	.747
Treatment * Ethnicity	25527.995	4	6381.999	6.994	.000
Year * Treatment * Ethnicity	1085.793	4	271.448	0.297	.880
Error	7311764.029	8013	912.488		
Total	867803368.000	8033			
Corrected Total	7670997.687	8032			

Note. * r^2 = .047 (Adjusted r^2 = .045).

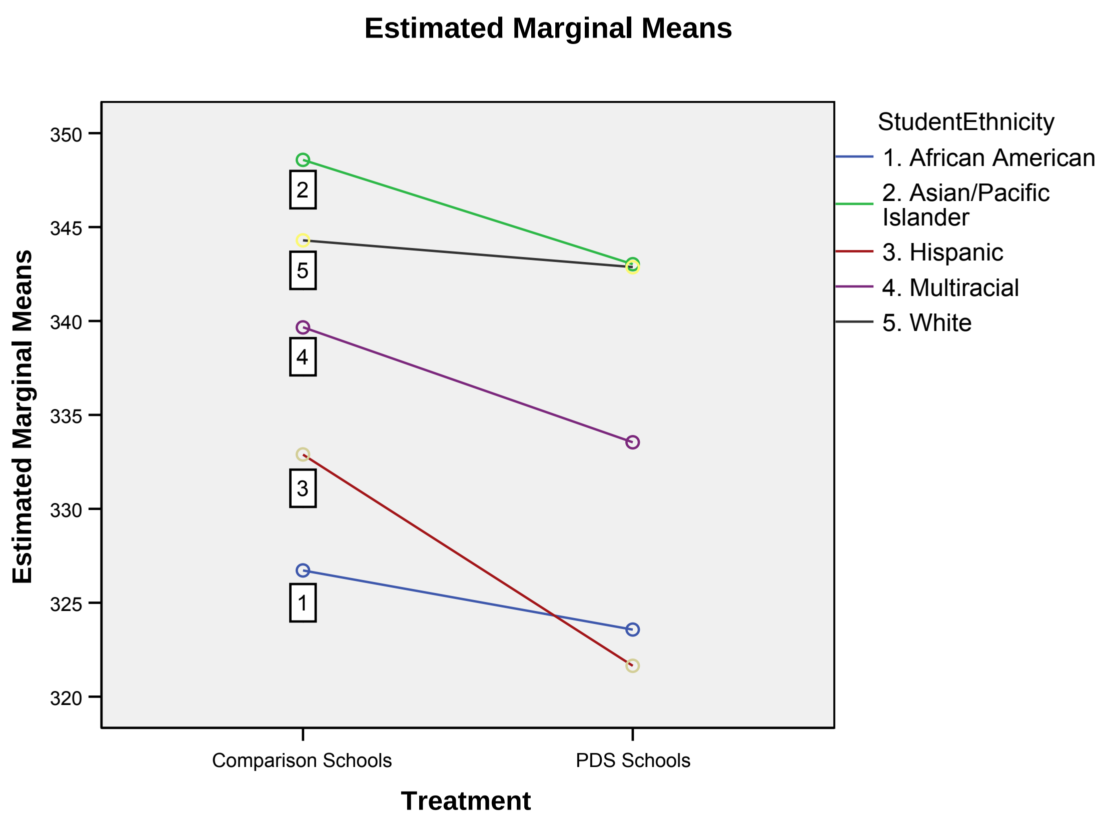

Figure 1. Profile Plot Comparing Estimated Marginal Means from Mathematics Scores of Elementary School Treatment Groups by Ethnic Group Subsets.

Table 3

Post-hoc Analysis Using Tukey B for Elementary School Students' Mathematics Scores

Ethnicity	N	Subset		
		1	2	3
Hispanic	2795	324.89		
African American	4294	325.37		
Multiracial	200		335.90	
White	237			343.75
Asian/Pacific Islander	507			344.66

Note. α = .05. Means for groups in homogenous subsets are displayed. Based on Type III Sum of Squares. Error term is Mean Square(Error) = 912.488. Uses Harmonic Mean Sample Size = 424.360. The group sizes are unequal. Type I error levels are not guaranteed.

Table 4

Tests of Between-Subjects Effects for Elementary School Students' Science Scores

Source	Type III Sum of Squares	df	Mean Square	F	p
Corrected Model	280438.373*	39	7190.728	14.267	.000
Intercept	30986636.850	1	30986636.850	61480.048	.000
Year	22.176	1	22.176	0.044	.834
Ethnicity	15523.644	4	3880.911	7.700	.000
Treatment	2466.894	1	2466.894	4.895	.027
Year * Ethnicity	1145.340	4	286.335	0.568	.686
Year * Treatment	16.638	1	16.638	0.033	.856
Ethnicity * Treatment	9903.793	4	2475.948	4.912	.001
Year * Ethnicity * Treatment	1583.578	4	395.894	0.785	.534
Error	2296275.316	4556	504.011		
Total	462824062.000	4596			
Corrected Total	2576713.689	4595			

* r^2 = .109 (Adjusted r^2 = .101).

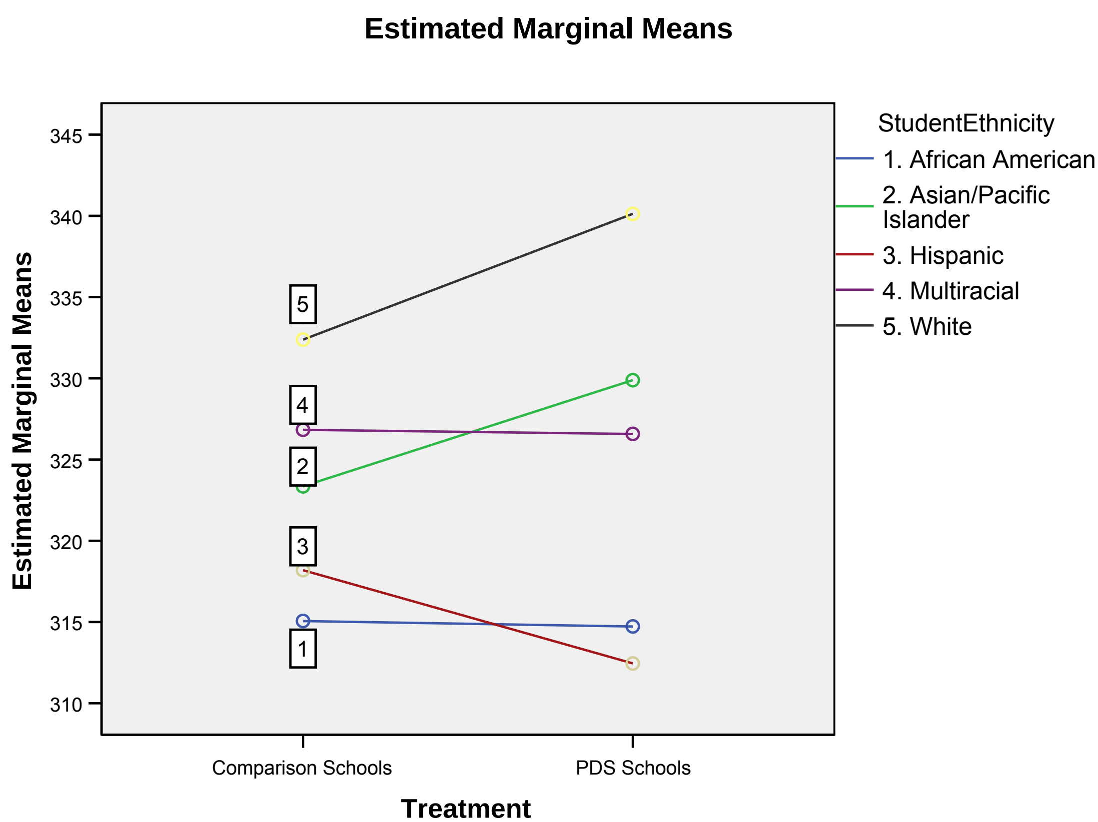

Figure 2. Profile Plot Comparing Estimated Marginal Means from Science Scores of Elementary School Treatment Groups by Ethnic Group Subsets.

interactions, no particular pattern was noted. Because the interactions do not address a research question, the interaction effects would not be considered practically significant.

The Tukey B test showed no significant difference in the means between Hispanic students and African American students (see Table 5). There was a difference between the combined subset of means of African American students and Hispanic students compared to the combined subset of means of Asian/Pacific Islander students and multiracial students, and the mean for White students was statistically significantly higher than either of the two other subsets.

Table 5

Post-hoc Analysis Using Tukey B for Elementary School Students' Science Scores

Ethnicity	N	Subset		
		1	2	3
Hispanic	1516	314.18		
African American	2511	314.83		
Multiracial	107		326.58	
Asian/Pacific Islander	305		327.86	
White	157			335.20

Note. α = .05. Means for groups in homogenous subsets are displayed. Based on Type III Sum of Squares. Error term is Mean Square(Error) = 504.011. Uses Harmonic Mean Sample Size = 249.354. The group sizes are unequal. Type I error levels are not guaranteed.

Middle School Mathematics

Middle school CRCT scores were collected from the PDS and comparison schools that participated in the project. The statistically significant differences identified by the F tests in the ANOVA for the middle schools with CRCT mathematics as the dependent variable were year, ethnicity, and the treatment by ethnicity interaction (see Table 6).

The statistical significance of ethnicity by treatment interaction is shown in Figure 3. The Tukey B test identified no significant difference in the means between African American students and Hispanic students (see Table 7). There was a difference between the combined subset of means of African American students and of means of Hispanic students and the combined subset of means of multiracial students and means of White students, as well as a difference between each of these subsets and the subset of means of Asian/Pacific Islander students; however, none of these was significant.

Table 6

Tests of Between-Subjects Effects for Middle School Students' Mathematics Scores

Source	Type III Sum of Squares	df	Mean Square	F	p
Corrected Model	819035.362*	19	43107.124	55.093	.000
Intercept	221397467.196	1	221397467.196	282954.171	.000
year	9175.882	1	9175.882	11.727	.001
Treatment	550.806	1	550.806	0.704	.401
Ethnicity	423662.605	4	105915.651	135.364	.000
year * Treatment	286.047	1	286.047	0.366	.545
year * Ethnicity	3460.354	4	865.088	1.106	.352
Treatment * Ethnicity	103076.557	4	25769.139	32.934	.000
year * Treatment * Ethnicity	277.998	4	69.499	0.089	.986
Error	8409771.687	10748	782.450		
Total	1055550977.000	10768			
Corrected Total	9228807.050	10767			

Note. * r^2 = .089 (Adjusted r^2 = .087).

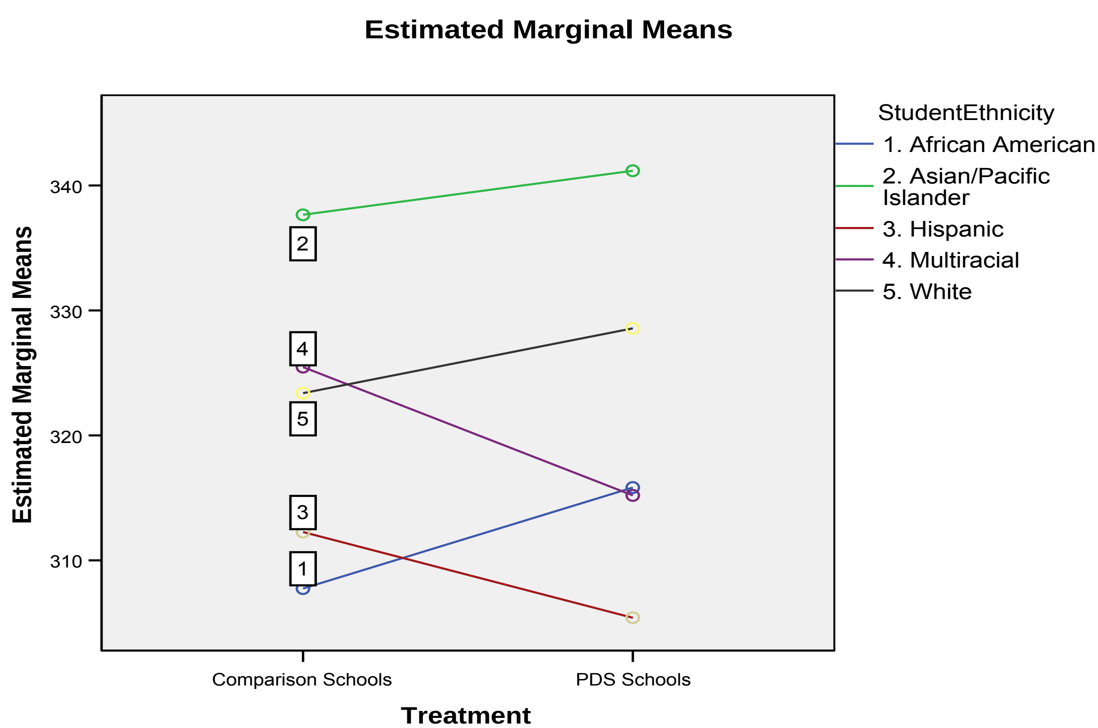

Figure 3. Profile Plot Comparing Estimated Marginal Means from Mathematics Scores of Middle School Treatment Groups by Ethnic Group Subsets.

Table 7

Post-hoc Analysis Using Tukey B for Middle School Students' Mathematics Scores

Ethnicity	N	Subset		
		1	2	3
Hispanic	2852	307.74		
African American	6720	309.88		
Multiracial	168		321.51	
White	431		324.46	
Asian/Pacific Islander	597			339.47

Note. α = .05. Means for groups in homogenous subsets are displayed. Based on Type III Sum of Squares. Error term is Mean Square(Error) = 782.450. Uses Harmonic Mean Sample Size = 478.604. The group sizes are unequal. Type I error levels are not guaranteed.

Middle School Science

Ethnicity and treatment by ethnicity were identified by the F tests in the ANOVA as statistically significant (see Table 8). Science scores were used as the dependent variable. The significance of the ethnicity by treatment interaction is shown in Figure 4. The Tukey B test identified four distinct, homogenous subsets of the mean scores associated with ethnicity: a subset of African American students, a subset of Hispanic students, a subset of multiracial students and White students, and a subset of Asian/Pacific Islander students (see Table 9). These differences were not significant.

High School Mathematics

The significant differences identified by the ANOVA with mathematics as the dependent variable were (a) year, (b) treatment, and (c) treatment by ethnicity (see Table 10). The significance of ethnicity by treatment interaction is shown in Figure 5. The Tukey B test showed no significant difference in the means between Hispanic students and African American students (see Table 11). There was a difference between

Table 8

Tests of Between-Subjects Effects for Middle School Students' Science Scores

Source	Type III Sum of Squares	df	Mean Square	F	p
Corrected Model	285485.526*	19	15025.554	34.605	.000
Intercept	134402872.898	1	134402872.898	309540.746	.000
Year	471.907	1	471.907	1.087	.297
Treatment	215.117	1	215.117	0.495	.482
Ethnicity	144215.678	4	36053.920	83.035	.000
Year * Treatment	14.983	1	14.983	0.035	.853
Year * Ethnicity	2593.642	4	648.411	1.493	.201
Treatment * Ethnicity	40096.183	4	10024.046	23.086	.000
Year * Treatment * Ethnicity	970.822	4	242.706	0.559	.692
Error	3126680.739	7201	434.201		
Total	701509861.000	7221			
Corrected Total	3412166.265	7220			

Note. * r^2 = .084 (Adjusted r^2 = .081).

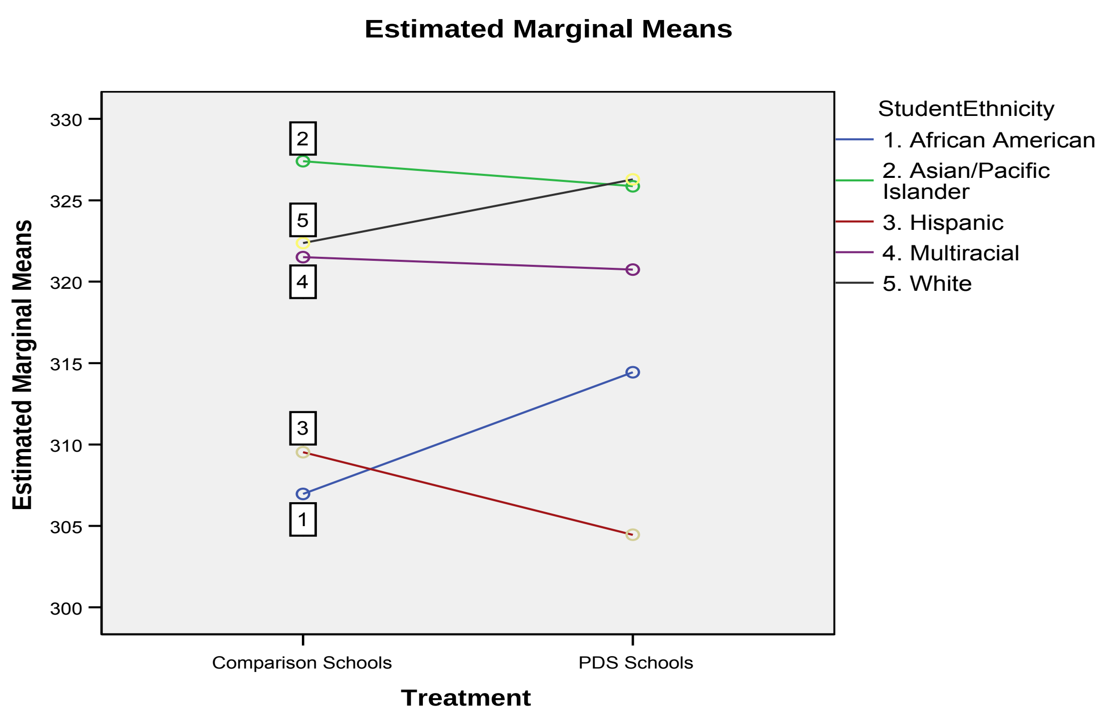

Figure 4. Profile Plot Comparing Estimated Marginal Means from Science Scores of Middle School Treatment Groups by Ethnic Group Subsets.

Table 9

Post-hoc Analysis Using Tukey B for Middle School Students' Science Scores

Student Ethnicity	N	Subset 1	2	3	4
Hispanic	2207	306.24			
African American	4050		310.34		
Multiracial	132			320.80	
White	344			323.30	323.30
Asian/Pacific Islander	488				325.59

Note. α = .05. Means for groups in homogenous subsets are displayed. Based on Type III Sum of Squares. Error term is Mean Square(Error) = 434.201. Uses Harmonic Mean Sample Size = 377.874. The group sizes are unequal. Type I error levels are not guaranteed.

Table 10

Tests of Between-Subjects Effects for High School Students' Mathematics Scores

Source	Type III Sum of Squares	df	Mean Square	F	P
Corrected Model	776373.315*	19	40861.753	59.725	.000
Intercept	333942064.329	1	333942064.329	488098.262	.000
Year	16776.774	1	16776.774	24.521	.000
Treatment	15455.273	1	15455.273	22.590	.000
Ethnicity	428364.197	4	107091.049	156.527	.000
Year * Treatment	256.692	1	256.692	0.375	.540
Year * Ethnicity	5191.642	4	1297.911	1.897	.108
Treatment * Ethnicity	25419.523	4	6354.881	9.288	.000
Year * Treatment * Ethnicity	3124.606	4	781.151	1.142	.335
Error	3403060.323	4974	684.170		
Total	1371101802.000	4994			
Corrected Total	4179433.638	4993			

Note. * r^2 = .186 (Adjusted r^2 = .183).

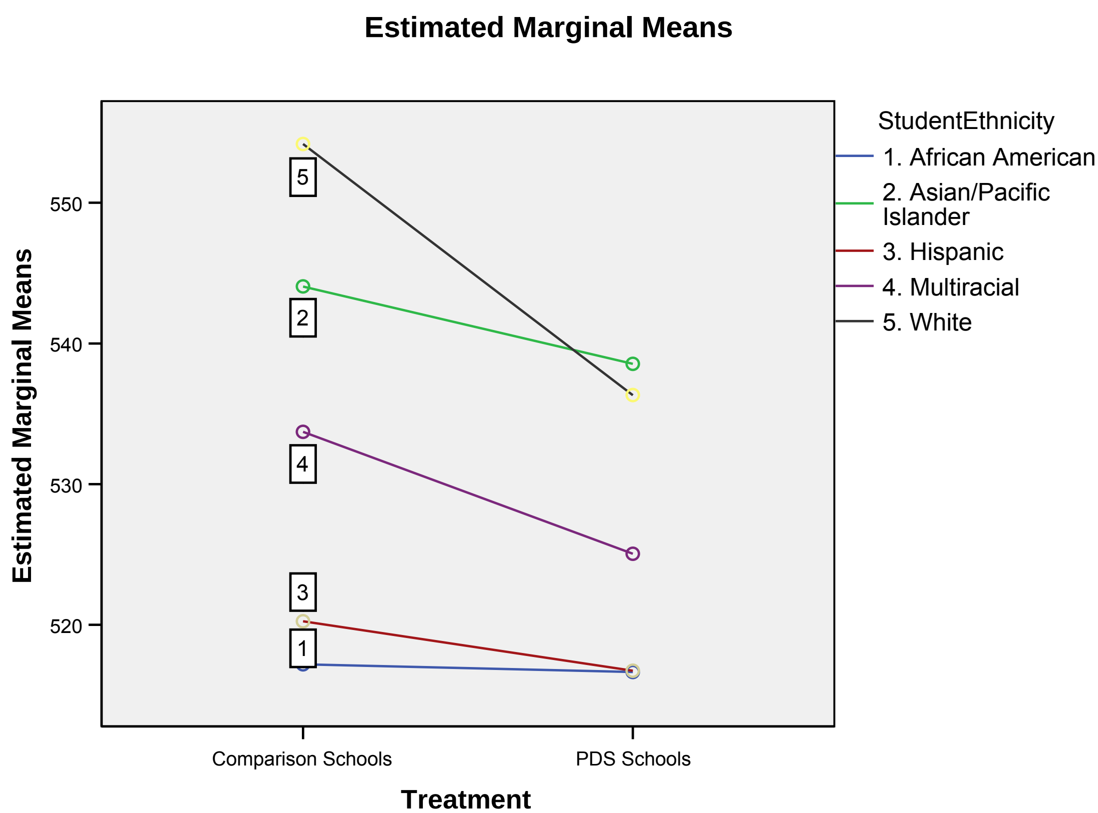

Figure 5. Profile Plot Comparing Estimated Marginal Means from Mathematics Scores of High School Treatment Groups by Ethnic Group Subsets.

Table 11

Post-hoc Analysis Using Tukey B for High School Students' Mathematics Scores

		Subset			
StudentEthnicity	N	1	2	3	4
African American	3299	517.14			
Hispanic	648	518.36			
Multiracial	80		530.43		
Asian/Pacific Islander	400			540.78	
White	567				550.35

Note. α = .05. Means for groups in homogenous subsets are displayed. Based on Type III Sum of Squares. Error term is Mean Square(Error) = 684.170. Uses Harmonic Mean Sample Size = 268.170. The group sizes are unequal. Type I error levels are not guaranteed.

African American/Hispanic means, Multiracial, Asian/Pacific Islander and White across all means but not at a significant level.

High School Science

Ethnicity and treatment by ethnicity were the significant differences identified by the ANOVA using science as the dependent variable (see Table 12). The significance of ethnicity by treatment interaction is shown in Figure 6. The Tukey B test identified four distinct, homogenous subsets of the mean scores associated with ethnicity: a subset of African American students, a subset of Hispanic students, a subset of Asian/Pacific Islander students and multiracial students, and a subset of White students (see Table 13). These differences were not significant.

School Assessment Using Adequate Yearly Progress

Adequate yearly progress by PDS and comparison schools changed very little over the 1-year period (see Figure 7). In 2004, 100% of the PDS elementary schools made AYP while 100% of the comparison elementary schools also made AYP. In 2005, the same held true. No elementary school was placed on the “needs status” list for 2006-2007. For middle schools in 2004, 75% made AYP for PDS while 0% made AYP for the comparison schools. In 2005, 75% of the PDS schools made AYP, but one school did not meet AYP that had the previous year and one school that did meet AYP the previous year did. For comparison schools 50% made AYP in 2005.Two PDS schools were on the needs status list, one for the 3rd year and one for the 6th year. If the school does not meet AYP during year 6, the school must be restructured with new administration and faculty hired. Four of the comparison schools were found on the needs status list for 2006-2007.

Table 12

Tests of Between-Subjects Effects for High School Students' Science Scores

Source	Type III Sum of Squares	df	Mean Square	F	p
Corrected Model	365328.678*	35	10437.962	24.119	.000
Intercept	31783292.686	1	31783292.686	73440.435	.000
Year	78.142	1	78.142	0.181	.671
Ethnicity	16964.030	4	4241.008	9.800	.000
Treatment	2859.257	1	2859.257	6.607	.010
Year * Ethnicity	2302.849	4	575.712	1.330	.256
Year * Treatment	7.131	1	7.131	0.016	.898
Ethnicity * Treatment	4138.015	4	1034.504	2.390	.049
Year * Ethnicity * Treatment	2255.106	4	563.777	1.303	.267
Error	1852283.311	4280	432.776		
Total	1079174580.000	4316			
Corrected Total	2217611.989	4315			

Note. * $r^2 = .165$ (Adjusted $r^2 = .158$).

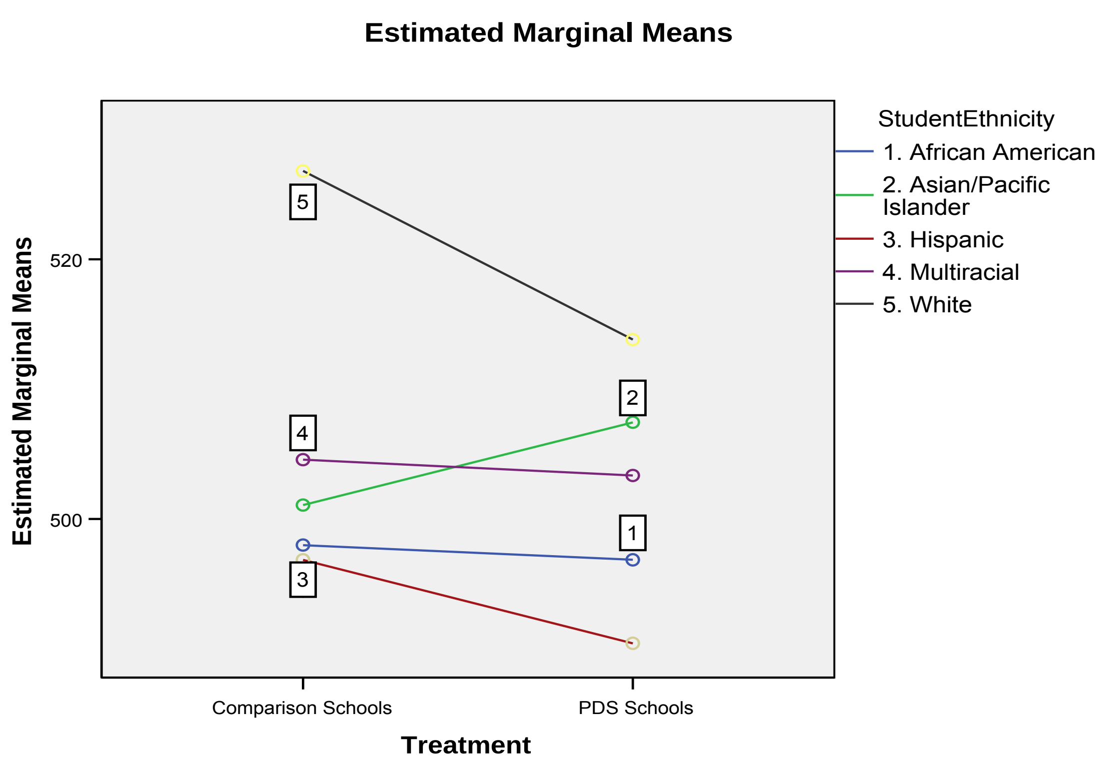

Figure 6. Profile Plot Comparing Estimated Marginal Means from Science Scores of High School Treatment Groups by Ethnic Group Subsets.

Table 13

Post-hoc Analysis Using Tukey B for High School Students' Science Scores

Ethnicity	N	Subset			
		1	2	3	4
Hispanic	466	491.58			
African American	3222		497.39		
Multiracial	55			504.09	
Asian/Pacific Islander	251			509.15	
White	322				524.11

Note. α = .05. Means for groups in homogenous subsets are displayed. Based on Type III Sum of Squares. Error term is Mean Square(Error) = 432.776. Uses Harmonic Mean Sample Size = 180.325. The group sizes are unequal. Type I error levels are not guaranteed.

One middle comparison school was in year 1, one was in Year 2, one was in Year 3, and one was in Year 6. Only one of the four PDS high schools met AYP in 2004. In 2005, again only 25% met AYP. The school that had previously made AYP did not in 2005 and one that did not in 2004 did make the AYP list in 2005. Comparison schools were similar to the PDS schools. in 2004, 25% made AYP and in 2005, again, 25% made AYP. The school that had previously made AYP did not in 2005, and one that did not in 2004 did make the AYP list For PDS high schools, two were on the needs status for Year 1, one for Year 2, and one for Year 3. In 2005, comparison schools had two in Year 1.

Correlating Change in Proportion Effect Sizes to Hedges's g Effect Sizes

In the previous section, the Adequate Yearly Progress of the PDS and comparison schools was discussed. One required aspect of AYP is based on the proportion of students at or above a specified cut score on the CRCT or HSGT. In this section, the change in proportions of students passing across the first year of the PDS implementation is examined in more detail. More specifically, the change in proportions passing using AYP standards is investigated to see the agreement, if any, with the change in scaled scores measuring academic achievement across the year.

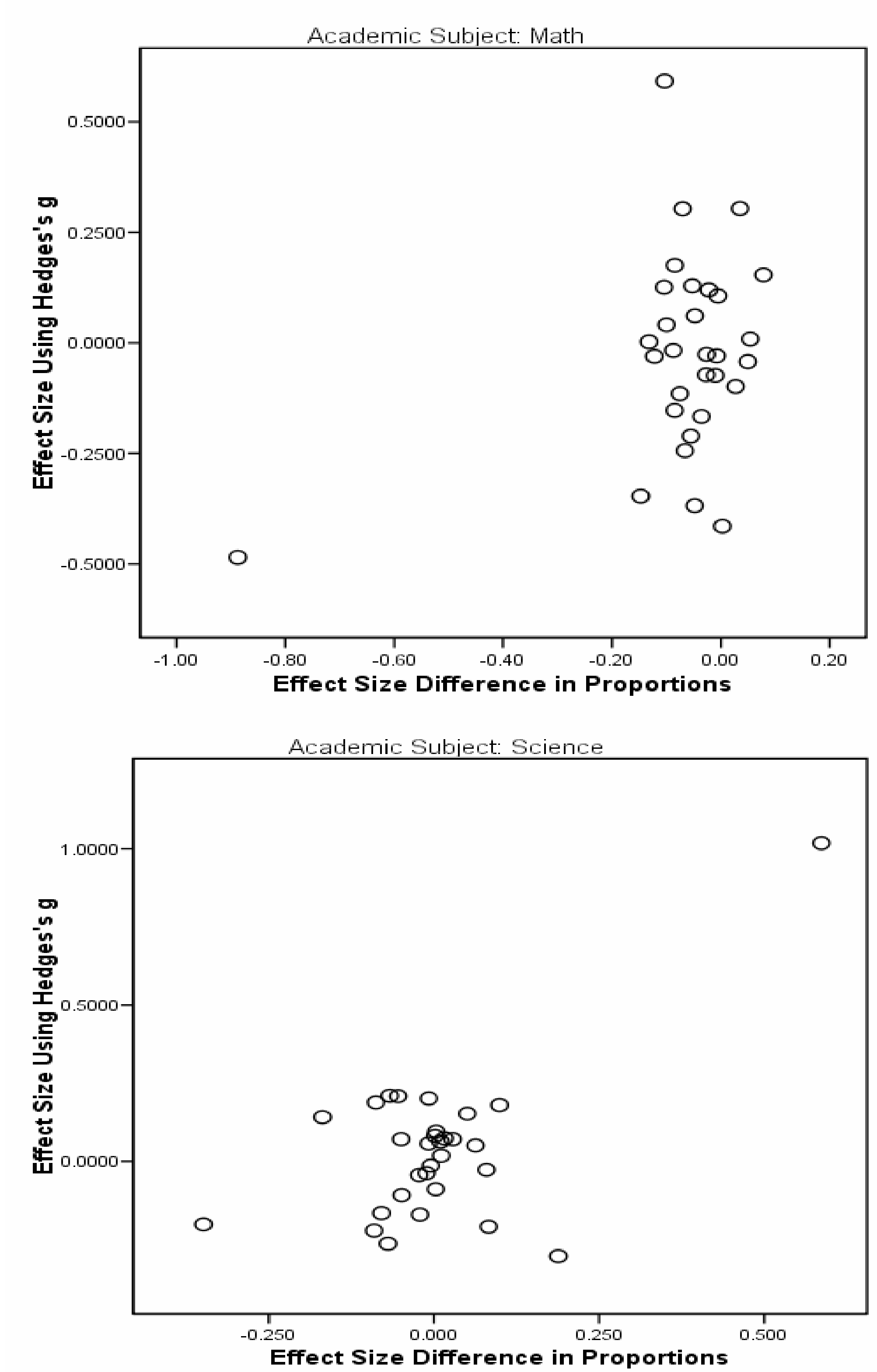

Figure 7. Ethnic Groups Scatterplots for Differences in Proportion Passing Across Baseline Compared to Year 1 and Scaled Score Effect Sizes for CRCT or HSGT. Each data point is the effect size measured as difference in proportions (Year 1 minus baseline year) and the effect size for scaled scores measured by Hedges's g (Year 1 minus baseline year).

Two different conceptual ways of measuring achievement change over a year are in terms of scale score changes and proportion changes based on a cut-score for passing. More specifically, for a given ethnic group, if the proportion passing increased from the baseline year to Year 1, it might be expected that the mean scale scores measuring achievement would increase from the baseline year to Year 1. This was the thought behind asking Research Question 6: From the baseline year CRCT and HSGT to the end of the first year within a given ethnic group, is there a correlation between the proportion passing change on the CRCT and HSGT for mathematics or science and Hedges's g effect size? The cut-score used for passing in this research question is the one used for Adequate Yearly Progress.

To investigate this research question for mathematics, I created a data set in the following manner. For all the elementary PDSs together, effect sizes were computed for change by subtracting baseline year from Year 1 in the effect size measures. There were five ethnic groups in the study, thus for the PDS elementary schools, there were five pairs of proportion type effect sizes and Hedges's g effect sizes. For the comparison elementary schools, there were also five pairs of proportion type effect sizes and Hedges's g effect sizes. In a similar manner, the middle schools would contribute 10 effect sizes and the high schools would contribute 10 effect sizes. Hence, in total, there were 30 pairs of effect sizes for mathematics. Using the same approach as when mathematics scores were the dependent variable, science scores as the dependent variable for ethnic groups resulted in another 30 effect sizes.

There are several ways of computing the effect sizes for the difference in proportions. One way is just to subtract the proportion passing at the end of the first year from the proportion passing at baseline year. Another way is to take the arcsine of the proportions in order to improve the scaling towards more equal units. When using just the difference in proportions, the numerical values for these pairs of effect sizes consisting of differences in proportions and Hedges's g are shown in Appendix B. The scatter plots of these pairs of effect sizes for mathematics and science are presented in Figure 7. The Pearson correlation for the complete data set ($n = 30$) for mathematics was .363 ($p < .005$). The Pearson correlation for the complete data set ($n = 30$) for science was .613 ($p < .001$).

These scatter plots suggested the presence of outliers in the data. (The outliers are identified in the tables in Appendix B.) In particular, the extreme points typically were with small groups of different students in a particular ethnic group across the two test administrations which could account for dramatic changes. When the outliers were removed, the correlations between the proportion effect sizes (measured as just the change in the proportions) and Hedges's g were not statistically significant. In particular, with the outlier removed in mathematics, the Pearson correlation was .028 ($n = 29$, $p = .885$). With the three outliers removed when science was the outcome variable, the Pearson correlation was .048 ($n = 27$, $p = .813$). It should be noted that if the outliers were left in the data, then the correlations would be statistically significant; thus, the decision to remove the outliers is of consequence.

Another effect size measure for changes in proportions defined as the arcsine of proportion 1 minus the arcsine of proportion 2 (Becker, 1994, p. 237), designated as b where $b = \arcsin(p_1) - \arcsin(p_2)$. After removing an outlier in the scatter plot, the Pearson correlation was run for the mathematics scores between the difference of the arcsine transformed proportions and Hedges's g effect sizes. The Pearson correlation was −.136 (n = 29, p = .482), which was not statistically significant.

In a similar fashion, after removal of outliers, the Pearson correlation was computed for the science scores between the difference of the arcsine transformed proportions and Hedges's g effect sizes. The Pearson correlation was −.044 (n = 27, p = .827), which did not obtain statistical significance. The data set used for this correlation is provided in Appendix B.

A different arcsine transformation for change in proportions is given by Cohen (1987, p. 181). More specifically, Cohen’s formula for an effect size index based on differences in proportions, designated by h, is

$$h = 2\arcsin\sqrt{p_1} - 2\arcsin\sqrt{p_2}\ .$$

Cohen states that this transformation, h, provides a solution to the issue of scale of equal units of detectability.

With the outliers removed, the Pearson correlation between Cohen’s h effect size for proportion change and Hedges’s g effect size for scaled scores for mathematics is −.128 (n = 29, p = .508). Removing the three outliers when science scores are the outcome variable, the Pearson correlation is −.023 (n = 27, p = .910).

Thus, using three different ways of computing effect size for the difference of proportions, the relationship between change in proportions over a year for different ethnic groups was not associated with the change in means over a year on the scaled scores. It was expected that there would be a positive correlation; hence, a finding of no correlation is a practically significant result. One implication of this finding is that a better understanding of student achievement would be given by reporting not only the proportion of an ethnic group which passed each year but also the mean achievement scaled scores for both the passing and not passing groups for each year.

Conclusions

Research Question 1

Research Question 1 addressed the effect of PDS^2 model on mean student achievement for mathematics and science across the study's year. Without regard to the comparison schools, six ANOVAs were run for professional development schools only. Three of the six ANOVAs were statistically significant: elementary schools with science as the dependent variable, middle school with mathematics as the dependent variable, high school with mathematics as the dependent variable (see Table 1).

Research Question 2

Research Question 2 dealt with the difference in achievement test score means between PDS^2 feeder pattern schools and comparison schools. There was no statistically significant year (baseline versus Year 1) by treatment interaction (professional development schools vs. comparison schools) in any of the six ANOVAs. However, there were some effects that were statistically significant. As shown in Table 2, in elementary school mathematics, there were significant between-subject effects with treatment, ethnicity, and treatment by ethnicity interaction. As shown in Table 4, in elementary school science, there were significant between-subject effects with treatment, ethnicity, and treatment by ethnicity interaction. As shown in Table 6, in middle school mathematics, there were significant between-subject effects with year, ethnicity, and treatment by ethnicity interaction. As shown in Table 8, in middle school science, there were significant between-subject effects with ethnicity and treatment by ethnicity interaction. As shown in Table 10, in high school mathematics, there were significant between-subject effects with year, treatment, ethnicity, and treatment by ethnicity interaction. As shown in Table 12, in high school science, there were significant between-subject effects with treatment, ethnicity, and treatment by ethnicity interaction. In contrast to Research Question 1, when the comparison schools are included and incorporated an ethnic group factor and interaction, there are no longer statistical gains from year 2005 to year 2006 (see Table 1). Thus, the PDS^2 model did not change mean student achievement in comparison to the comparison schools over the initial year of implementation.

Research Question 3

Research Question 3 addressed the Adequately Yearly Progress status of both PDSs and comparison schools when the baseline year is compared to the end of the first year. Adequate Yearly Progress status did not significantly change between the baseline year and Year 1 (see Table 14). While several schools did meet Adequately Yearly Progress, several others did not.

Research Question 4

Research Question 4 addressed the existence of the achievement gap in 24 schools averaged across 2 years of testing. These data provide contextual information for the study. All six ANOVAs, when analyzing the data for mean difference between ethnic groups, were statistically significant collapsing across year (baseline vs. Year 1) and treatment (PDS vs. comparison school; see Tables 2, 4, 6, 8, 10, 12).

Table 14

AYP Status Over Year

Treatment Category	Educational Group	AYP 2004-05	AYP 2005-06	Needs Status 2006-07
PDS	Elementary	Met	Met	n/a
PDS	Elementary	Met	Met	n/a
PDS	Elementary	Met	Met	n/a
PDS	Elementary	Met	Met	n/a
PDS	Middle	Met	Met	n/a
PDS	Middle	Not Met	Met	Year 6
PDS	Middle	Met	Met	n/a
PDS	Middle	Met	Not Met	Year 3
PDS	High	Met	Not Met	Year 1
PDS	High	Not Met	Not Met	Year 3
PDS	High	Not Met	Not Met	Year 1
PDS	High	Not Met	Met	Year 2
Comparison	Elementary	Met	Met	n/a
Comparison	Elementary	Met	Met	n/a
Comparison	Elementary	Met	Met	n/a
Comparison	Elementary	Met	Met	n/a
Comparison	Middle	Not Met	Met	Year 6
Comparison	Middle	Not Met	Not Met	Year 1
Comparison	Middle	Not Met	Met	Year 3
Comparison	Middle	Not Met	Not Met	Year 2
Comparison	High	Not Met	Met	n/a
Comparison	High	Not Met	Not Met	Year 1
Comparison	High	Not Met	Not Met	n/a
Comparison	High	Met	Not Met	Year 1

Research Question 5

Research Question 5 addressed the issue of closing the gap. As shown in Chapter 4, there were no year-by-ethnicity interactions that were statistically significant. Hence, I concluded that there is no evidence that the gap did not close during this 1-year period.

Research Question 6

As previously discussed in this chapter, after removing outliers, all correlations between the change over 1 year in proportion passing with the change over 1 year in scaled scores were not statistically significant (see Appendix C). Until it is shown that there is a high correlation between change in proportion passing and change in CRCT scaled scores for different ethnic groups, there should be dual reporting of both proportion passing and scaled scores for achievement. This finding suggests a reporting weakness in *No Child Left Behind Act of 2001*.

CHAPTER 5

DISCUSSION

Research Questions

Six research questions guided my investigation of student achievement in mathematics and science during the first year of implementation of the professional development school model at 12 urban schools at the elementary school, middle school, and high school levels.

1. How does the PDS^2 model affect mean student achievement in mathematics and science as measured by the Georgia Criterion Referenced Competency Tests and Georgia High School Graduation Tests?
2. Are there significant differences in mean achievement test scores between PDS^2 feeder pattern schools and comparison schools?
3. From the baseline year to the end of the first year, how many PDSs and comparison schools have changed their Adequate Yearly Progress status and in what direction?
4. Is there a mean difference between ethnic groups on the scaled scores of the CRCT and HSGT for mathematics and science?
5. From the baseline year CRCT and HSGT to the end of the first year by ethnic groups, is there a mean difference in scaled scores on the CRCT and HSGT for mathematics or science?

6. From the baseline year CRCT and HSGT to the end of the first year for different ethnic groups, is there a correlation between the proportion passing change on the CRCT and HSGT for mathematics or science and Hedges's g effect size?

Discussion

The first two research questions are concerned with the effect of the PDS model on student achievement between PDS baseline and Year 1 and between PDS and comparison schools. In the PDS^2 research project, individual achievement test scores were collected at the elementary, middle, and high school levels in both PDS and comparison schools. A quasi-experimental research model was used with comparison schools matched to PDS schools. Student achievement in mathematics and science, used as the dependent variable, was the primary research method used to make a decision about the efficacy of Year 1 PDS implementation.

Collection and analysis of achievement test scores have also been an important part of previously conducted PDS research studies. In three major PDS research studies, significant difference in achievement test scores were often difficult to obtain. The Benedum Collaborative Model of Teacher Education (Hoffman et al., 1997, p. 36) included standardized student achievement test scores as a portion of the research model. The Stanford Achievement Test-9 (SAT-9) scores were collected at the elementary, middle, and high school levels. It was noted that most of the PDS schools participating in the project lacked baseline data. The researcher compared achievement test scores with state and county averages to determine an increase or decrease in scores. Statistically significant scores were found in grades 6, 7, and 8 with the largest gains being made in mathematics. Ultimately, Van Dempsey decided it was more important to examine the magnitude of the difference through the use of effect sizes in addition to calculating statistical significance. This decision was made because while some test scores may have been of a substantial magnitude they did not achieve statistical significance. There was a similar occurrence in the PDS^2 research model. Although this study PDS^2 did include baseline data instead of relying on state and county averages, it was difficult to obtain statistically significant differences in the achievement test scores. Calculation of effect sizes and confidence intervals were included in the original research design in an effort to allow me to examine the magnitude of the difference.

Along with the quantitative analysis of student achievement scores, the Benedum Collaborative Model of Teacher Education (Hoffman et al., 1997) also included student, intern, and teacher interviews along with collaborative partner interviews. Qualitative interview methods were included in an effort to get at the less quantifiable professional development school variables. An effort was made in the research model to tease out less quantifiable strands that directly affected student achievement. No consideration was given for the level of development of the PDS school. All PDSs were included, those newly formed as well as PDSs that were several years old.

A second large-scale PDS research study, the Houston Consortium, collected achievement test data in 16 elementary PDS schools. The data were collected over a baseline plus 2-year period and included qualitative data collection in the research design. Achievement test scores were collected, and student time-on-task and teacher observations were made. The research results showed that teachers in PDS schools had increased test scores and that they were also more encouraging and responsive to students' academic needs. The PDS research project used an elementary, middle, and high school feeder pattern spanning all grade levels. The elementary research model was the cleanest with one teacher assigned to one specific class. This model is not available to middle and high schools because students have multiple teacher assignments. The multiple teacher assignments make teacher–student–test score linkage much more complicated and confounding. High School PDSes are the most difficult to research primarily because of the use of multiple teachers for individual students. The Houston Consortium chose to use only elementary schools for their PDS research. In fact, the majority of PDS research conducted has taken place in elementary schools. The Consortium also looked at data over a baseline plus 2-year implementation period during which there was significant change in academic achievement test scores. The PDS^2 model chose to use an elementary, middle, high school feeder pattern that makes teacher-student-test scores linkage much more difficult. I suspect that the degree of teacher participation in the PDS model can make a difference in academic achievement. This would be particularly true at the middle and high school level. The Houston Consortium analyzed data over multiple years. The PDS^2 research only looked at Year 1 implementation. I believe that multiple years of data collection would give the best opportunity for identifying standardized student achievement gain.

The third major study was the State of Maryland PDS research study. This study used only PDS schools that were in their fourth year of implementation or older. The idea of using more mature PDS schools was to collect data from PDS schools that had a fully implemented PDS model over several years with teachers dedicated to the mission of the PDS. State standardized test scores were collected over a 7-year period and analyzed for trends. PDS test scores were compared to state test result data. The Maryland PDS research also included action research in an effort to understand the teaching and learning process better. Recommendations that came from this research included the use of alternative assessments, journals, and report cards and the use of culturally relevant practices, strategies, and content. It was recommended that additional qualitative research methods be used to capture the intangible educational improvements inherent in the PDS model. The PDS^2 research is focused on studying schools that are in the beginning phase of PDS implementation. Administration and faculty must be in agreement on PDS model implementation before student achievement impact can be seen. I would suspect that the increase in academic achievement test score improvement would increase as the PDS school begins to develop and mature over time.

Although there are major differences in the research design of the four major research projects being discussed, several research design recommendations can be made. It seems that using a quasi-experimental design would be useful to help justify the academic achievement students could potentially make in a fully developed PDS school. Also, using PDS schools that have been established for at least 4 years levels the research playing field. It is not reasonable to expect a significant difference in standardized test scores with PDS schools that are in their initial development. Implementation of a PDS plan is complicated and requires total administration and faculty buy in. Previous research and PDS^2 research has shown that it is difficult to show a difference in standardized achievement test schools during the first several years of implementing a PDS model.

In this dissertation, I delimited myself by choosing to use only quantitative data for analysis. In the three previous research studies, qualitative components were included in the research design. Teitel (2003b) speaks to the intrinsically qualitative nature of the professional development school movement because of the many unquantifiable variables associated with education in general and PDS in particular. Including the use of student, intern, and teacher interviews, parent and faculty focus groups, journals, report cards and student and teacher portfolios would provide additional information that would help in isolating PDS effects from other confounding variable. Many researchers are in agreement that it is difficult to account accurately for educational variables that are intertwined, making them difficult to identify. As researchers begin to identify these variables, it will be easier to determine if particular PDS programs are successes or failures.

Coupling the use of student achievement data with qualitative data analysis is the best chance the PDS movement has to show that an educational difference can be made. Continuous systematic PDS research is important in order to assess the effectiveness of the PDS implementation. Perhaps even more important is an agreement on what the minimum requirements are for PDS implementation so that the movement can become more standardized. While most PDS researchers do not agree with standardization, without some agreement evaluation of the PDS program is difficult. Inability to show that the PDS model is effective may negatively affect the PDS movement.

Adequate Yearly Progress status did not significantly change between the baseline year and Year 1. While different PDS and comparison schools met AYP and did not meet AYP, the numbers remained the same. Of greater concern are the schools that did not make AYP going into the 6th year. Failure to make AYP during the 2006-07 year will trigger a complete school restructuring from the administration to the faculty. Two middle schools fall into this category. One is a PDS^2 school and one is a comparison school. The other schools have more time to meet the AYP standards before possible restructuring.

From the data collected for the PDS^2 grant, there were significant mean main effects collapsing over years and treatment implementation in all six ANOVAS. The data showed that in many cases, White students and Asian students had higher mean achievement than Hispanic students and African American students. One discovery was that while academic gaps were identified between ethnicities and while there were minimal closure of these gaps in some instances, the differences were not statistically significant except for middle and high school mathematics when collapsed over years. Collapsing across ethnic groups showed no statistically significant difference of means for mathematics or science in elementary school.

This data set also showed that there was no statistically significant closure of the achievement gap. With the current conversation around the closure of the achievement gap between ethnicities, this is a practically significant and important finding. Based on these findings, researchers can see that it is important to continue to analyze achievement gap data by ethnicity to increase the likelihood of making good policy choices and producing beneficial educational practice. Longitudinally collected PDS data are beginning to document the positive and/or negative effects of PDS on student achievement.

One surprising research discovery was the relationship between the change in proportion passing across the year and the change in scale scores across the year. In other words, two different ways of measuring achievement change over a year are in terms of scale score changes and proportion changes based on the cut score for passing. With outliers removed, it was found that there was essentially no correlation between these two measures of academic achievement. This has educational policy implications. In looking at the data, it appears that student scores falling near the cut score may have been pushed over to passing. The implication is that these students may have received special treatment because they were identified as the easiest to bump up to passing. When these students' scores rise above the cut-off score, this causes the mean achievement for students passing to decrease. It would actually appear that student learning had decreased when in fact the number of students attaining the cut score had not. As the pass rate is inflated, those that are failing are failing with significantly lower scores, causing, I believe, a different, potentially worse teaching and learning situation for teachers the next year. It could be inferred that based on current AYP policy teachers are now teaching to the students who fall on the "bubble" and are very close to making the AYP cut score. This could leave students who achieve at much higher levels unchallenged and students who are failing significantly educationally underserved.

Because of the near zero correlation between the scale score changes and proportion passing changes, I would tentatively call for additional research in this area so that prevailing educational policy concerning AYP pass rates could be studied and revised if needed. One piece of evidence that this research has produced is that in the case of these student scores, it should be asked why there is a near zero correlation between these two indicators of academic achievement.

In holistically reviewing the total PDS^2 data collected, I found that there was no overall significant difference attributable to the PDS^2 intervention. One possible reason that this occurred is that reform at the school level seldom occurs meaningfully in a short period (Southern Regional Education Board, 2006), although some changes may be observable. Previous PDS studies have not shown significant change during the first year of implementation of a project. In fact, the majority of large PDS research studies use a minimum of 2 years and up to 7 years when comparing data. The Maryland PDS project collected data over 7 years in an effort to give a more stable picture of the PDS implementation. Also, including more qualitative methods such as focus groups, and teacher and student interviews and portfolios will help provide a more holistic view of PDS^2.

One assumption which appeared to be supported was that the science and mathematics tests between years were reasonably equated, although this was not specifically tested. When this research was initially conceptualized, I planned to look at all test scores included on the CRCT and HSGT. That plan was soon changed because the Georgia Department of Education made the decision to change the tests to in an effort to realign the tests with the new Georgia Performance Standards. State officials verified that only mathematics and science could be used because of the realignment of the tests and because a calibration statistic had not been published. Still several of the participating schools were given new mathematics and/or science tests based on the new performance standards, causing a problem with the original statistical method to be used. Originally three MANOVAS were to be run; however, because of the change in tests even within mathematics and science, this was not possible. Often students had mathematics tests that had not been changed but had science tests that had been changed, and the reverse was true in some cases. The missing science scores at some grade levels resulted in the use of six ANOVAs for statistical analysis rather than three MANOVAs, which would have had considerable missing data.

Another assumption was that the data provided by the school systems were accurate and as requested. Because the data set was large, it required a great deal of expertise by the school system data manager. The assumption is that the data manager reported the data accurately as requested for all students and subjects. A great deal of time was spent preparing the data for analysis; however, the ultimate quality of the data set rests with the system reporting the data.

Limitations

Several limitations of the study were identified. First, it is possible that the PDS^2 intervention dosage was low. Because the schools were in the initial phase of becoming a PDS, it is safe to assume that not all of the PDS interventions written into the PDS^2 grant were implemented. In fact, several of the interventions were not implemented until well into Year 2. Another factor could be that analyzing data at the student level and reporting it at the system level does not give the true picture of the PDS^2 intervention. I would suggest that a better way to evaluate PDSs would be to do so at the student and teacher level of specific classrooms. These classrooms would be identified as PDS classrooms with more than one intern placed within the class and all participating in a biweekly professional learning community. There would be more control of the PDS^2 intervention and easier data collection.

Second was the use of an untreated comparison group design with separate pretest and posttest samples. This design is the most frequently used in the social sciences but is difficult to draw causal conclusions because of internal threats to validity.

The third limitation is related to the use of student achievement as the dependent variable. Because identifying specific educational strands (independent variables) and activities related to them are difficult, identifying direct causation links from variable to outcome is virtually impossible.

Implications

The findings for Research Question 6 have policy implications. With outliers removed from the data, the finding of no correlation between changes in proportions passing across a year with changes in scaled score achievement across a year has policy implications for reporting Adequate Yearly Progress. In particular, if the proportion passing increases, it is possible for the mean scale score achievement to decrease. Thus, a much better understanding of the students' passing and their achievement is given by reporting both the proportion passing for a year and the mean scaled score achievement for both the students' passing and not passing. In fact, not reporting the data fully may lead to less discussion of important academic issues. Currently the Georgia Department of Education only reports pass rates based on predetermined cut scores. I believe the current type of reporting encourages teachers to teach to the students just under the passing cut score so that pass rates will improve, potentially leaving gifted students and failing students academically underserved. This method gives no recognition or credit to teachers of students or students who exceed standards above 350. Conversely, no credit is given to the teachers of students or students who increase failing scores significantly but not to passing. Reporting both scores would give parents, teachers, and students a clearer understanding of the academic achievement made during the time period being tested. The results of my research suggest that reporting of both the proportion passing for a year and the mean scaled score should be instituted so that there is a better understanding of academic achievement for each student until there is a high statistical correlation between the two scores.

Future Research

Because it is difficult to obtain differences in standardized achievement test scores for systems and schools over a period of 1-2 years, future studies should include the use of qualitative research methods. The inclusion of qualitative research methods would give additional ways to identify intertwined educational strands that affect the academic achievement of students within a PDS setting. Considering the inclusion of student, intern, and teacher interviews, parent and faculty focus groups, student and intern journals, report card data, and student, intern, and teacher portfolios would give additional important information on the efficacy of the PDS^2 program and its impact on student academic achievement.

After conducting the research, I also found that consideration should also be given to the level of development of the PDS. Most studies that obtain a significant difference in standardized achievement test scores are conducted with schools that have been a PDS for 4 years or longer. Collecting data from schools that are in the first few years of implementation does not give a meaningful picture of the impact of the PDS model.

Many new PDS programs require a minimum of 2 years for initial implementation of the program for there to begin to be a difference in the program delivery at the school level. The fact that my data were collected from the first year of implementation gives a clear snapshot of baseline data collected but does not give a true picture of the potential impact of the program on student achievement.

Difficulties have also been encountered in the assignment of university faculty to specific school sites. It is much easier to identify and assign an Early Childhood Education university faculty member to an elementary school because of their multiple subject expertise. In my research, I found that it was much easier to link academic achievement data to elementary teachers and university faculty than in middle and high schools where this type of linking becomes more complex as the number of teachers and subjects per student grows. Also, with the PDS^2 model, only one university faculty member was assigned to each middle and high school, and that faculty member had expertise in only one subject area. This left a huge gap in the delivery of professional development services in areas outside the university faculty's expertise. In the future, the use of teams of university faculty with multiple areas of expertise should be considered for middle and high schools.

Also under discussion is the possibility of researching how the PDS^2 model affects several classrooms within one school as opposed to the entire school or system. The proposed Teacher-Intern-Professor (TIP) approach places several interns in two or more classrooms with a professor using the PDS^2 model. The cooperating teachers and the university faculty members are matched based on content area and all participate in a bi-weekly learning community. Both quantitative and qualitative data would be collected to document the use of the model and student achievement. The TIP model would allow for a more focused PDS^2 implementation and documentation of student academic achievement.

Three of the six ANOVAs showed significant change in achievement means for the PDS schools when using the PDS school data only. However, when data from both PDS schools and matched comparison schools were analyzed, the overall results indicated no statistically significant gains in mathematics and science means for the professional development schools in relation to the comparison schools for the first year of professional development school implementation. While the study did not show statistically significant gains, it does provide baseline data for other potential research projects as well as an opportunity for data driven adjustments in the implementation and/or model of the grant. There are several additional studies that could come from the data collected for this dissertation but were beyond the scope of this work. Additional data were collected on gender, ESOL, and special education.

Finally, an outcome of this study shows that in the future, continuous systematic PDS data collection over time will be the best test of the efficacy of the program. I would encourage additional studies over the life of the PDS^2 grant so that a solid research base could be established for future PDS researchers. Future researchers should be able to use this project to build upon when seeking to conduct similar analysis of professional development schools or schools implementing similar reform.

REFERENCES

Abdal-Haqq, I. (1996). An information provider's perspective on the professional development movement. *Contemporary Education, 67*(4), 237-239.

Abdal-Haqq, I. (1998). *Professional development schools: Weighing the evidence.* Thousand Oaks, CA: Corwin Press.

Abdal-Haqq, I. (1993). *Resources on professional development schools: An annotated bibliography.* Washington, DC: ERIC Clearinghouse on Teacher Education and the American Association for College of Teacher Education.

American Psychological Association. (1994). Publication manual of the American Psychological Association (4th ed.). Washington, DC: Author.

Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care*, 42, 1-16.

Bainbridge, W.L., & Lasley, T.J. (2002). Demographics, diversity, and K-12 accountability: The Challenge of closing the achievement gap. Education and Urban Society, 34(4), 422-437.

Becker, B. J. (1994). Combining significance levels. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 215-230). New York: Russell Sage Foundation.

Book, C. L. (1996). Professional development schools. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 194-210). New York: Macmillan.

Bowman, K.L. (2001). The new face of school desegregation. *Duke Law Journal, 50*(6), 1751-1808.

Campoy, R. W. (2000). *A professional development school partnership: Conflict and collaboration*. Westport, CT: Bergin & Garvey.

Carnegie Foundation for the Advancement of Teaching. (1920). *15th annual report of the President and Treasurer*. New York: Author.

Clark, R. W. (1999). *Effective professional development schools*. San Francisco: Jossey-Bass.

Clark, R. W. (1997). *Professional development schools: Policy and financing*. Washington, DC: AACTE.

Clark, R. W. (2000). School-university partnerships and professional development schools. *Peabody Journal of Education, 74,* 164-177.

Cohen, E. G. (1981). Sociology looks at team teaching. *Research in Sociology of Education and Socialization, 2,* 163-193.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities.* New York: Basic Books.

Cook, T. D. (2002). Randomized experiments in education: Why are they so rare? *Educational Evaluation and Policy Analysis, 24*(3), 175-199.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation design & analysis issues for field settings*. Boston: Houghton Mifflin Company.

Cooper, D. H., & Corbin, C. T.(2003). Student Achievement in Maryland's Professional Development Schools. In D.L. Wiseman & S. L. Knight (Eds.), *Linking school-university and K-12 student outcomes* (pp. 71-86). New York: AACTE Publications.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Darling-Hammond, L. (1992). Teaching and knowledge: Policy issues posed by alternate certification for teachers. *Peabody Journal of Education, 67*(3), 123-154.

Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappan, 78*(3), 193-202.

Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education, 51*(3), 166-173.

Darling-Hammond, L. (2005). *Professional Development Schools: Schools for developing a profession.* New York: Teachers College Press.

Darling-Hammond, L. (2006). Constructing 21st century teacher education. *Journal of Teacher Education, 57*(3), 300-314.

Demie, F. (2001). Ethnic and gender differences in educational achievement and implications for school improvement strategies. *Educational Research, 43*(1), 91-106.

Dewey, J. (1896). The university school. *University Record, 5,* 417-442.

Dixon, P. L., & Ishler, R. (1992). Acceptable approaches to alternative certification. *Teacher Education and Practice, 8*(1), 29-35.

Dobson, D., & Cook, T. J. (1980). Avoiding type III errors in program evaluation: Results from a field experiment. *Evaluation and Program Planning, 3,* 269-376.

Fenstermacher, G. (1992). The place of alternate certification in the education of teachers. *Peabody Journal of Education, 67*(3), 155-185.

Flexner, A. (1914). *Medical education in the United States and Canada: A report to the Carnegie Foundation for the Advancement of Teaching.* New York: Carnegie Foundation for the Advancement of Teaching.

Fraser, J. W. (1992). Preparing teachers for democratic schools: The Holmes and Carnegie reports five years later: A critical reflection. *Teachers College Record, 94*(1), 7-40.

Frasier, C. M. (1994). Two education reform trains: Standards/assessment and simultaneous renewal. *Record in Educational Leadership, 14*(2), 15-17.

Freire, P. (1970). *Pedagogy of the oppressed*. New York. The Seabury Press.

Frey, N. (2002). Literacy achievement in an urban middle-level professional development school: A learning community at work. *Reading Improvement. 39*(1), 3-13.

Fuhrman, S. (1999). *The new accountability*. Philadelphia: Consortium for Policy Research in Education.

Fulton, K, Yoon, I, & Lee, C. (2005). Induction into learning communities. National Commission on Teaching and America's Future, August.

Garcia, E. (2001). *Understanding and meeting the challenge of student diversity* (3rd ed.). Boston: Houghton Mifflin.

Garcia, E. E., & Cuellar, D. (2006). Who are these linguistically and culturally diverse students? *Teachers College Record, 108*(11), 2220-2246.

Georgia Department of Education. (n.d.a). *Criterion-Referenced Competency Test (CRCT).* Retrieved November 11, 2006, from http://www.gadoe.org/-ci_testing.aspx?PageReq=CI_TESTING_CRCT

Georgia Department of Education. (n.d.b). *Georgia High School Graduation Tests (GHSGT).* Retrieved November 9, 2006, from http://public.doe.k12.ga.us/-ci_testing.aspx?PageReq=CI_TESTING_GHSGT

Georgia Department of Education. (n.d.c). *Online assessment system.* Retrieved November 11, 2006, from http://www.doe.k12.ga.us/curriculum/testing/ghsgt.asp

Georgia Department of Education. (2006). *Curriculum frequently asked questions.* Retrieved November 9, 2006, from http://www.georgiastandards.org/faqs.aspx

Georgia Partnership for Excellence in Education. (2006). Gap analysis: Georgia's unfinished business in teacher quality. Retrieved December 04, 2006, from http://www/gpee/org

Gill, B., & Hove, A. (2000). *The Benedum collaborative model of teacher education: A preliminary study.* Pittsburg: Rand Education.

Gliessman, D., Pugh, R., Dowden, D., & Hutchins, T. (1988). Variables influencing the acquisition of a generic teaching skill. *Review of Educational Research, 38*(1), 25-46.

Goldrick-Rab, S., & Mazzeo, C. (2005). What no child left behind means for college access. *Review of Research in Education, 29*, 107-129.

Goodlad, J. I. (1980). How laboratory schools go awry. *UCLA Educator, 21*(2), 46-53.

Goodlad, J. I. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.

Goodlad, J. I. (1990). *Teachers for our nation's schools*. San Francisco: Jossey-Bass.

Goodlad, J. I. (1994). The National Network for Educational Renewal. *Phi Delta Kappan, 75*(8), 632-639.

GreatSchools Inc. (n.d.). *No Child Left Behind (NCLB) requirements for schools.* Retrieved November 11, 2006, from http://www.greatschools.net/definitions/-nclb/nclb.html

Green, S. B., & Salkind, N.J. (2005). *Using SPSS for windows and macintosh: Analyzing and understanding data, fourth ed.* Upper Saddle River, NJ: Pearson-Prentice Hall.

Gresham, F., MacMillan, D. L., Beebe-Frankenberger, M. B., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning disabilities research and practice, 15,* 198-205.

Griffith, J. (2002). A multilevel analysis of the relation of school learning and social environments to minority achievement in public elementary schools. *The Elementary School Journal,* 102(5), 349-365.

Grissmer, D., & Kirby, S. N. (1987). *Teacher attrition: The uphill climb to staff the nation's schools.* Santa Monica, CA: The RAND Corporation, R-3512-CSTP.

Grusky, D. B. (1994). The contours of social stratification. In D. B. Grusky (Ed.), Social stratification in sociological perspective: Class, race and gender pp. 3-35. Boulder: Westview.

Guarino, C.M., Santibanez, L., Daley, G. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, 76(2), 173-208.

Gutierrez, F.D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19-25.

Hair, Jr., J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate data analysis with readings.* New York: Macmillan Publishing.

Hallinan, M. T., & Khmelkov, V. T. (2001). Recent developments in teacher education in the United States of American. *Journal of Education for Teaching, 27*(2), 175-185.

Hanushek, E.A. (2004). Our school performance matters. *Journal of Education, 185*(3), 1-6.

Hedges, L, & Oklin, I. (1985). *Statistical methods for meta-analysis*. Boston, MA: Harcourt, Brace Jovanovich, Publishers.

Hoffer, T., Greeley, A. M., & Coleman, J. S. (1985). Achievement growth in public and catholic schools. *Sociology of Education*, 58, 74-97.

Hoffman, N. E., Reed, W. M., & Rosenbluth, G. S. (1997). Introduction. In N. E. Hoffman, W. M. Reed, & G. S. Rosenbluth (Eds.), *Lessons from restructuring experiences: Stories of change in professional development schools* (pp. 1-6). Albany: State University of New York Press.

Holmes Group. (1986). *Tomorrow's teachers: A report of the Homes Group.* East Lansing, MI: Author.

Holmes Group. (1990). *State of principles: Tomorrow's schools: Principles for the design of professional development schools.* East Lansing, MI: Author.

Holmes Group. (1995). *Tomorrow's schools of education.* East Lansing, MI: Author.

Houston, W. R., Clay, D., Hollis, L. Y., Ligons, C., Roff, L., & Lopez, N. (1995). *Strength through diversity: Houston consortium for professional development and technology centers.* (ERIC Document Reproduction Service No. ED 404303)

Houston, W. R., Hollis, L. Y., Clay, D., Ligons, C. M., & Roff, L. (1999). Effects of collaboration on urban teacher education programs and professional development schools. In D. M. Byrd & J. McIntyre (Eds.), *Research on Professional Development Schools: Teacher education yearbook VII* (pp. _____). Thousand Oaks, CA: Corwin Press, Inc.

Imig, D., & Imig, S. (2005). The learned report on teacher education: a vision delayed. Change Magazine, September/October, electronic copy.

Judge, H., Carriendo, R., & Johnson, S. M. (1995). *Professional Development Schools and MSU: The report of the 1995 review.* Michigan State University College of Education. Retrieved August 30, 2006, from http://ed-web3.educ.msu.edu/pds/-archive.html

Johnson, S. & Birkeland, S. (2003). Pursuing a "sense of success": New teachers explain their career decisions. *American Educational Research Journal*, 40(3), 581-617.

Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746-759.

Knight, S. L., Wiseman, D. L., & Cooner, D. (2000). Using collaborative teacher research to determine the impact of professional development school activities on elementary students' math and writing outcomes. *Journal of Teacher Education, 51*(1), 26-38.

Lee, C.D. (2003). Why we need to re-think race and ethnicity in educational research. *Educational Researcher*, 32(5), 3-5.

Levin, M. (2002). Why invest in professional development schools? *Educational Leadership, 59*(6), 65-70.

Levin, M. (Ed.). (1992). *Professional practice schools: Linking teacher education and school reform.* Teachers College Press: New York.

Levin, M., & Churins, E. J. (1999). Designing standards that empower professional development schools. *Peabody Journal of Education, 74*(3&4), 178-208.

Levin, M., & Trachtman, R. (1997). *Building professional development schools: Politics, practice and policy.* New York: Teachers College Press.

Lutonsky, L. (1972). *Portal schools*. Washington, DC: Council of the Great City Schools.

McDonnell, L. (2004). *Politics, persuasion and educational testing.* Cambridge, MA: Harvard University Press.

Medley, D. M. (1977). *Teacher competence and teacher effectiveness: A review of process product research.* Washington, DC: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED 143 629)

Murry, F. B. (1993). "All or none" criteria for professional development schools. In P. G. Alback, H. G. Petrie, M.J. Shujaa, & L. Weis (Eds.), *Educational policy: Volume 7, Number 1. Professional development schools* (pp. 61-73). Newbury Park, CA: Corwin Press.

National Center for Education Statistics (NCES) (2002). Digest of education statistics. Retrieved August 28, 2006 from DATABSE.

National Council for Accreditation of Teacher Education. (1997). *Draft standards for identifying & supporting quality professional development schools.* Washington, DC: Author.

National Council for Accreditation of Teacher Education. (2001). *Standards for professional development schools.* Washington, DC: Author.

National Council on Teaching & America's Future. (1996). *What matters most: Teaching for America's future.* New York: Author.

Neapolitan, J. E., & Scott, C. (2004). The NCATE professional development school standards field test project. In J. E. Neapolitan, T. D. Proffitt, C. L. Wittmann, & T. R. Berkeley (Eds.), *Traditions, standards, & transformations: A model for professional development school networks* (pp. 145-155). New York: Peter Lang Publishing, Inc.

No Child Left Behind Act. 20 U.S.C. § 6301. (2001).

Osguthorpe, R. T., Harris, R. C., Harris, M. F., & Black, S. (1995). Appendix: National Network for Educational Renewal site descriptions. In R. T. Osguthorpe, R. C. Harris, J. F. Harris, & S. Black (Eds.), *Partner schools: Centers for educational renewal* (pp. 283-310). San Francisco: Jossey-Bass.

Petersilia, J. (1990). Conditions that permit intensive supervision. *Crime and Delinquencey, 36*(1). 126-145.

Petrie, H. G. (Ed.). (1995). *Professionalization, partnership, and power: Building professional development schools.* Albany: State University of New York Press.

Pine, G. J. (2003). Making a differences: A professional development school's impact on student learning. In D. L. Wiseman & S. L. Knight (Eds.), *Linking school-university collaboration and K-12 student outcomes.* Washington, DC: AACTE Publications.

Powell, M. (1978). Research on teaching. *Educational Forum, 43,* 27-37.

Pugach, M. C., & Johnson, L. J. (2002). *Collaborative Practitioners, Collaborative Schools.* Love Publishing Company: Denver, CO.

Reed, C. J., Kochan, F. K., Ross, M. E., & Kunkel, R. C. (2001). Designing evaluation systems to inform, reform, and transform professional development schools. *Journal of Curriculum and Supervision, 16*(3) 188-205.

Research Forum on Children, Families, and the New Federalism (2002). Lack of appropriate research leads to gaps in knowledge about children in immigrant families. *Forum* 5(1):1.

Riehl, C. (2006). Feeling better: A comparison of medical research and education research. *Educational Researcher, 35*(5), 24-29.

Rivkin, E, Hanushek, A. & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2) 417-458.

Robinson, S., & Darling-Hammond, L. (1994). Change for collaboration and collaboration for change: Transforming teaching through school-university partnerships. In L. Darling-Hammond (Ed.), *Professional development schools: Schools for developing a profession* (pp. 203-219). New York: Teachers College Press.

Ross, D., Brownell, M., Sindelar, P., & Vandiver, F. (1999). Research from professional development schools: Can we live up to the potential? *Peabody Journal of Education, 74*(3&4), 209-223.

Rossi, P. H. & Freeman, H. E. (1989). *Evaluation: A systematic approach. 4th eds*. Newbury Park, CA: Sage.

Rice, E. H. (2002). The collaboration process in professional development schools: Results of a meta-ethnography, 1990-1998. *Journal of Teacher Education, 53*(1), 55-67.

Rosenshine, B. (1976). Classroom instruction. In N.L. Gage (Ed.), *The psychology of teaching methods* (77th yearbook of the National Society for the Study of Education, Part 1, pp. 335-371). Chicago: University of Chicago Press.

Scales, P.C., Roehlkepartain, E.C., Neal, M., Kielsmeier, J.C. & Benson, P.L. (2006). Reducing academic achievement gaps: The role of community service and service-learning. *Journal of Experiential Education, 29*(1), 38-60.

Schutt, R. K. (1996). *Investigating the social world.* Thousand Oaks, CA: Pine Forge Press.

Sewell, T. E., Shapiro, J. P., Ducett, J. P., & Sanford, J. S. (1995). Professional development schools in the inner city: Policy implications for school-university collaboration. In H. G. Petrie (Ed.), *Professionalization, partnership, and power: Building professional development schools* (p. 179-198). Albany: State University of New York Press.

Shea, C. (1996). Psychologists debate accuracy of significance test. *Chronicle of Higher Education, 42*(49), A12, A16.

Slavin, R. E. (2004). Education research can and must address "what works" questions. *Educational Researcher, 33*(1), 27-28.

Snyder, J. (1999). Professional development schools: What? So what? Now what? *Peabody Journal of Education*, 74(3&4), 136-144.

Snyder, J., Morrison, G., & Smith, R. C. (1996). *Dare to dream: Educational guidance for excellence.* Indianapolis, IN: Lilly Endowment.

Southern Regional Education Board. (2006). *Establishing benchmarks for progress for middle grades sites*. Atlanta, GA: Author.

Sowards, W. (1969). Florida State University model for the preparation of elementary school teachers. *Journal of Research and Development in Education, 2*(3), 22-30.

SPSS Inc. (2005). SPSS 14.0 for windows. USA.

Shadish, W., Cook, T., Campbell, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin Company.

Stallings, J. A., & Kowalski, T. (1990). Research on professional development schools. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 251-263). New York: Macmillan.

Stevens, J.P. (2002). Applied multivariate statistics for the social sciences. Mahwah, NJ: Lawrence Earlbaum Associates, Inc.

Su, Z. (1999). Creating an equal partnership for urban school and teacher education renewal. In D. M. Byrd & D. J. McIntyre (Ed.), *Research on professional development schools* (pp. 46-60). Thousand Oaks, CA: Corwin Press, Inc.

Sykes, G. (1997). Worthy of the name: Standards for professional development schools. In M. Levine & R. Trachtman (Eds.), *Making professional development schools work: Politics, practices, and policy* (pp. 159-184). New York: Teachers College Press.

Teitel, L. (1999). Looking toward the future by understanding the past: The historical context of professional development schools. *Peabody Journal of Education, 74*(3 & 4), 6-20.

Teitel, L. (with Abdal-Haqq, I). (2000). Assessing the impacts of professional development schools. In I. Abdal-Haqq (Series Ed.), *The Professional Development Practice Series* [Booklet]. Washington: AACTE.

Teitel, L. (2001a). *How professional development schools make a difference: A review of research.* Washington, DC: National Council for Accreditation of Teacher Education.

Teitel, L. (2001b). An assessment framework for professional development schools. *Journal of Teacher Education, 52*(1), 57-71.

Teitel, L. (2003a). *The professional development schools handbook.* Thousand Oaks, CA: Corwin Press, Inc.

Teitel, L. (2003b). *How professional development schools make a difference: A review of research* (2nd ed., rev.). Washington: NCATE.

The National Commission on Excellence in Education. (1992). *A Nation at Risk: The full account.* Portland, OR: USA Research, Inc.

Thompson, B. (2002). “Statistical, “practical,” and “clinical”: How many kinds of significance do counselors need to consider? *Journal of Counseling & Development, 80*(winter), 64-71.

Thompson, B. (2006). *Foundations of Behavioral Statistics*. New York: The Guilford Press.

Tracey, T.J. (2000). Issues in the analysis and interpretation of quantitative data: Deinstitutionalization of the null hypothesis test. In S. D. Brown & R. Lent (Eds.) *Handbook of counseling psychology* (3rd ed., pp. 177-198). New York: Wiley.

Trubowitz, S., & Longo, P. (1997). *How it works inside a school-college collaboration.* New York: Teachers College Press.

Vacha-Haase, T. & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*(4) 473-481.

Valli, L., Cooper, D., & Frankes, L. (1997). Professional development schools and equity: A critical analysis of rhetoric and research. In M. W. Apple (Ed.), *Review of Research in Education, 22*(251-304). Washington, DC: American Educational Research Association.

Webb-Dempsey, J. (1997). Reconsidering assessment to be reflective of school reform. In N. E. Hoffman, W. M. Reed, & G. S. Rosenbluth (Eds.), *Lesson from* *restructuring experiences: Stories of change in professional development school* (pp. 269-294). Albany: State University of New York Press.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604. [reprint available through the APA Home Page: http://www.apa.org/journals/amp/amp548594.html]

Williams, R. (1993). A professional development school initiative: Indiana State University story. *Contemporary Education, 64*(4), 210-214.

Wiseman, D. L., & Cooner, D. (1996). Discovering the power of collaboration. *Teacher Education and Practice, 12*(1), 18-28.

Wright, D. P., McKibbon, M., & Walton, P. (1987). *The effectiveness of the teacher trainee program: An alternate route into teaching in California.* Sacremento, CA: California Commission on Teacher Credentialing.

Zimpher, N. (1990). Creating professional development school sites. *Theory Into Practice, 29,* 42-49.

APPENDIXES

APPENDIX A

Description of Baseline and Year 1 PDS Participants' Scores in Mathematics

Level	Ethnicity	N_{Year1}	$N_{baseline}$	M_{Year1}	$M_{baseline}$	SD_{Year1}	$SD_{baseline}$
Elem	Total	2180	2279	325.13	324.90	29.031	30.824
Elem	1	871	1013	323.14	324.00	28.081	30.212
Elem	2	174	186	341.59	344.46	27.500	30.360
Elem	3	1013	978	322.51	320.77	27.500	29.248
Elem	4	65	58	333.58	333.50	27.960	36.245
Elem	5	57	44	342.07	343.66	39.252	34.465
Middle	Total	2196	3557	315.95	312.35	27.871	30.818
Middle	1	1113	1913	317.33	312.50	25.877	28.460
Middle	2	158	271	343.23	339.01	28.241	36.190
Middle	3	825	1191	307.67	304.15	26.303	27.951
Middle	4	27	49	321.74	316.69	30.698	33.500
Middle	5	73	133	327.11	327.70	27.496	37.579
High	Total	1145	1143	522.19	517.83	26.713	22.661
High	1	743	769	518.02	515.24	24.688	21.633
High	2	116	119	542.40	534.71	27.553	22.753
High	3	215	177	520.34	513.12	25.980	20.772
High	4	20	20	525.60	524.50	24.641	27.912
High	5	51	58	543.41	529.24	27.046	20.464

Effect sizes compare baseline to Year 1 for each ethnic group:

Level	Ethnicity	Cohen's d	Hedges's g	Confidence Interval for Hedges's g	
Elem	Total	0.007677	0.007675	-0.051040	0.066394
Elem	1	-0.029410	-0.029390	-0.119970	0.061181
Elem	2	-0.098920	-0.098710	-0.305560	0.108129
Elem	3	0.061328	0.061305	-0.026580	0.149191
Elem	4	0.002490	0.002474	-0.351550	0.356503
Elem	5	-0.042690	-0.042360	-0.435730	0.351008
Middle	Total	0.121099	0.121083	0.067846	0.174321
Middle	1	0.175392	0.175348	0.101326	0.249370
Middle	2	0.126018	0.125796	-0.070570	0.322164
Middle	3	0.128991	0.128943	0.040073	0.217813
Middle	4	0.155179	0.153601	-0.316800	0.624003
Middle	5	-0.017170	-0.017110	-0.302610	0.268394
High	Total	0.176007	0.175949	0.093839	0.258059
High	1	0.119907	0.119848	0.018931	0.220765
High	2	0.304718	0.303736	0.046532	0.560939
High	3	0.303726	0.303142	0.103086	0.503197
High	4	0.041782	0.040952	-0.578920	0.660823
High	5	0.596188	0.591999	0.207635	0.976364

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White.

Description of Baseline and Year 1 PDS Participants' Scores in Science

Level	Ethnicity	N_{Year1}	$N_{baseline}$	M_{Year1}	$M_{baseline}$	SD_{Year1}	$SD_{baseline}$
Elem	Total	1204	1303	317.22	314.51	22.506	23.548
Elem	1	502	611	317.01	312.45	20.594	22.686
Elem	2	102	109	330.76	329.02	24.495	24.582
Elem	3	543	521	313.06	311.83	21.732	21.834
Elem	4	27	34	328.33	324.82	21.623	26.621
Elem	5	30	28	339.93	340.32	27.062	27.174
Middle	Total	1031	2732	312.31	311.85	23.610	22.342
Middle	1	535	1118	314.63	314.25	23.466	20.577
Middle	2	79	270	326.80	324.93	25.380	25.554
Middle	3	370	1167	304.06	304.84	20.474	20.374
Middle	4	13	49	321.38	320.10	28.829	24.044
Middle	5	34	128	328.41	324.16	18.032	23.503
High	Total	983	1161	497.56	497.12	19.527	24.964
High	1	711	807	497.93	495.78	18.694	25.441
High	2	59	109	504.97	509.93	23.142	23.741
High	3	176	173	491.11	489.73	18.626	20.340
High	4	15	17	500.00	506.71	21.514	21.523
High	5	22	55	515.91	511.73	21.743	19.122

Effect sizes compare baseline to Year 1 for each ethnic group:

Level	Ethnicity	Cohen's d	Hedges's g	Confidence Interval for Hedges's g	
Elem	Total	0.117553	0.117518	0.039099	0.195937
Elem	1	0.209486	0.209345	0.090957	0.327732
Elem	2	0.070905	0.070650	-0.199450	0.340747
Elem	3	0.056469	0.056429	-0.063800	0.176654
Elem	4	0.143007	0.141181	-0.364680	0.647044
Elem	5	-0.014380	-0.014190	-0.529220	0.500845
Middle	Total	0.020268	0.020264	-0.051380	0.091905
Middle	1	0.017630	0.017622	-0.085420	0.120661
Middle	2	0.073290	0.073132	-0.177640	0.323901
Middle	3	-0.038240	-0.038220	-0.155170	0.078726
Middle	4	0.051049	0.050408	-0.561140	0.661952
Middle	5	0.189025	0.188137	-0.190570	0.566846
High	Total	0.019440	0.019433	-0.065520	0.104387
High	1	0.095411	0.095364	-0.005510	0.196235
High	2	-0.210760	-0.209810	-0.527390	0.107773
High	3	0.070789	0.070636	-0.139270	0.280542
High	4	-0.311820	-0.303960	-1.002260	0.394345
High	5	0.210148	0.208040	-0.287490	0.703565

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White.

Description of PDS and Comparison Participants' Baseline Scores in Mathematics

Level	Ethnicity	N_{PDS}	$N_{Comparison}$	M_{PDS}	$M_{Comparison}$	SD_{PDS}	$SD_{Comparison}$
Elem	Total	2279	1760	324.90	329.12	30.824	32.882
Elem	1	1013	1175	324.00	325.57	30.212	32.252
Elem	2	186	74	344.46	348.11	30.360	39.360
Elem	3	978	397	320.77	331.82	29.248	30.656
Elem	4	58	39	333.50	339.79	36.245	31.314
Elem	5	44	75	343.66	346.09	34.465	34.430
Middle	Total	3557	3065	312.35	307.38	30.818	29.171
Middle	1	1913	2252	312.50	304.02	28.460	27.965
Middle	2	271	106	339.01	335.42	36.190	30.628
Middle	3	1191	519	304.15	311.38	27.951	27.764
Middle	4	49	48	316.69	319.19	33.500	31.909
Middle	5	133	140	327.70	321.33	37.579	32.879
High	Total	1143	1261	517.83	522.75	22.661	34.162
High	1	769	797	515.24	513.58	21.633	34.312
High	2	119	83	534.71	538.95	22.753	26.676
High	3	177	123	513.12	518.28	20.772	21.568
High	4	20	12	524.50	528.50	27.912	15.377
High	5	58	246	529.24	548.95	20.464	25.158

Effect sizes compare PDS to comparison group for each ethnic group:

Level	Ethnicity	Cohen's d	Hedges's g	Confidence Interval for Hedges's g	
Elem	Total	-0.132970	-0.132940	-0.195210	-0.070680
Elem	1	-0.050120	-0.050100	-0.134150	0.033943
Elem	2	-0.110090	-0.109770	-0.379320	0.159781
Elem	3	-0.372540	-0.372340	-0.489810	-0.254870
Elem	4	-0.183070	-0.181630	-0.588310	0.225057
Elem	5	-0.070550	-0.070100	-0.442400	0.302205
Middle	Total	0.165298	0.165279	0.116892	0.213666
Middle	1	0.300779	0.300725	0.239441	0.362009
Middle	2	0.103391	0.103184	-0.121470	0.327842
Middle	3	-0.259190	-0.259080	-0.362530	-0.155620
Middle	4	-0.076400	-0.075800	-0.473970	0.322385
Middle	5	0.180726	0.180226	-0.057580	0.418034
High	Total	-0.168130	-0.168080	-0.248260	-0.087890
High	1	0.057654	0.057627	-0.041470	0.156721
High	2	-0.173500	-0.172850	-0.453650	0.107953
High	3	-0.244530	-0.243920	-0.474820	-0.013010
High	4	-0.166070	-0.161890	-0.878680	0.554901
High	5	-0.809730	-0.807720	-1.100930	-0.514510

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White.

Description of PDS and Comparison Participants' Baseline Scores in Science

Level	Ethnicity	N_{PDS}	$N_{Comparison}$	M_{PDS}	$M_{Comparison}$	SD_{PDS}	$SD_{Comparison}$
Elem	Total	1303	1039	314.51	315.18	23.548	26.089
Elem	1	611	690	312.45	313.08	22.686	24.402
Elem	2	109	47	329.02	320.68	24.582	38.345
Elem	3	521	223	311.83	315.11	21.834	27.910
Elem	4	34	22	324.82	326.55	26.621	16.964
Elem	5	28	57	340.32	331.91	27.174	21.688
Middle	Total	2732	2505	311.85	310.39	22.342	20.332
Middle	1	1118	1701	314.25	308.40	20.577	18.661
Middle	2	270	105	324.93	325.00	25.554	24.168
Middle	3	1167	515	304.84	310.21	20.374	21.532
Middle	4	49	47	320.10	320.57	24.044	20.716
Middle	5	128	137	324.16	321.12	23.503	23.405
High	Total	1161	1253	497.12	505.40	24.964	23.818
High	1	807	825	495.78	499.77	25.441	20.602
High	2	109	77	509.93	512.81	23.741	23.173
High	3	173	105	489.73	494.53	20.340	20.779
High	4	17	13	506.71	506.54	21.523	21.647
High	5	55	233	511.73	527.71	19.122	21.822

Effect sizes compare PDS to comparison group for each ethnic group:

Level	Ethnicity	Cohen's d	Hedges's g	Confidence Interval for Hedges's g	
Elem	Total	-0.027117	-0.027109	-0.1086334	0.0544160
Elem	1	-0.026682	-0.026666	-0.1355514	0.0822189
Elem	2	0.283902	0.282517	-0.0609403	0.6259744
Elem	3	-0.137729	-0.137589	-0.2945902	0.0194117
Elem	4	-0.074105	-0.073071	-0.6095316	0.4633888
Elem	5	0.356161	0.352933	-0.1024910	0.8083577
Middle	Total	0.068211	0.068201	0.0139664	0.1224363
Middle	1	0.300874	0.300794	0.2249244	0.3766639
Middle	2	-0.002781	-0.002775	-0.2281964	0.2226466
Middle	3	-0.258980	-0.258865	-0.3629215	-0.1548080
Middle	4	-0.020910	-0.020743	-0.4209239	0.3794380
Middle	5	0.129624	0.129254	-0.1119394	0.3704482
High	Total	-0.339680	-0.339575	-0.4199894	-0.2591597
High	1	-0.172565	-0.172486	-0.2697063	-0.0752651
High	2	-0.122511	-0.122011	-0.4140534	0.1700311
High	3	-0.234072	-0.233435	-0.4766819	0.0098115
High	4	0.007879	0.007666	-0.7144751	0.7298073
High	5	-0.748885	-0.746919	-1.0470122	-0.4468267

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White.

Description of PDS and Comparison Participants' Year 1 Scores in Mathematics

Level	Ethnicity	N_{PDS}	$N_{Comparison}$	M_{PDS}	$M_{Comparison}$	SD_{PDS}	$SD_{Comparison}$
Elem	Total	2180	1814	325.13	330.82	29.031	30.771
Elem	1	871	1235	323.14	327.86	28.081	31.188
Elem	2	174	73	341.59	349.04	27.500	31.509
Elem	3	1013	407	322.51	333.96	27.500	27.360
Elem	4	65	38	333.58	339.53	27.960	26.568
Elem	5	57	61	342.07	342.49	39.252	32.806
Middle	Total	2196	1950	315.95	312.63	27.871	27.174
Middle	1	1113	1442	317.33	309.80	25.877	26.573
Middle	2	158	62	343.23	338.81	28.241	27.031
Middle	3	825	317	307.67	315.50	26.303	25.654
Middle	4	27	44	321.74	329.27	30.698	25.053
Middle	5	73	85	327.11	322.28	27.496	27.897
High	Total	1145	1445	522.19	528.56	26.713	29.190
High	1	743	990	518.02	520.80	24.688	25.184
High	2	116	82	542.40	549.15	27.553	28.477
High	3	215	133	520.34	522.21	25.980	25.267
High	4	20	28	525.60	538.93	24.641	22.996
High	5	51	212	543.41	559.41	27.046	25.239

Effect sizes compare PDS to comparison group for each ethnic group:

Level	Ethnicity	Cohen's d	Hedges's g	Confidence Interval for Hedges's g	
Elem	Total	-0.190723	-0.190687	-0.2531168	-0.1282579
Elem	1	-0.157636	-0.157580	-0.2444351	-0.0707247
Elem	2	-0.259254	-0.258460	-0.5327269	0.0158068
Elem	3	-0.416970	-0.416750	-0.5327931	-0.3007064
Elem	4	-0.216693	-0.215080	-0.6164010	0.1862420
Elem	5	-0.011647	-0.011571	-0.3726469	0.3495046
Middle	Total	0.120528	0.120507	0.0594644	0.1815487
Middle	1	0.286616	0.286531	0.2079352	0.3651277
Middle	2	0.158379	0.157834	-0.1362628	0.4519301
Middle	3	-0.299716	-0.299519	-0.4296189	-0.1694190
Middle	4	-0.275648	-0.272641	-0.7538913	0.2086087
Middle	5	0.174289	0.173449	-0.1398966	0.4867954
High	Total	-0.226513	-0.226448	-0.3042401	-0.1486549
High	1	-0.111322	-0.111274	-0.2064816	-0.0160660
High	2	-0.241602	-0.240676	-0.5244503	0.0430986
High	3	-0.072733	-0.072576	-0.2888653	0.1437139
High	4	-0.562701	-0.553477	-1.1378895	0.0309365
High	5	-0.625121	-0.623323	-0.9336192	-0.3130264

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White.

Description of PDS and Comparison Participants' Year 1 Scores in Science

Level	Ethnicity	N_{PDS}	$N_{Comparison}$	M_{PDS}	$M_{Comparison}$	SD_{PDS}	$SD_{Comparison}$
Elem	Total	1204	1050	317.22	319.24	22.506	22.337
Elem	1	502	708	317.01	317.05	20.594	23.342
Elem	2	102	47	330.76	326.02	24.495	21.228
Elem	3	543	229	313.06	321.27	21.732	17.684
Elem	4	27	24	328.33	327.13	21.623	23.716
Elem	5	30	42	339.93	332.86	27.062	20.637
Middle	Total	1031	953	312.31	308.19	23.610	21.190
Middle	1	535	696	314.63	305.52	23.466	19.509
Middle	2	79	34	326.80	329.79	25.380	21.459
Middle	3	370	155	304.06	308.83	20.474	23.210
Middle	4	13	23	321.38	322.43	28.829	20.090
Middle	5	34	45	328.41	323.62	18.032	21.137
High	Total	983	919	497.56	496.65	19.527	19.489
High	1	711	879	497.93	496.20	18.694	19.157
High	2	59	*	504.97	*	23.142	*
High	3	176	12	491.11	499.17	18.626	20.657
High	4	15	10	500.00	502.60	21.514	14.439
High	5	22	12	515.91	525.92	21.743	25.618

Effect sizes compare PDS to comparison group for each ethnic group:

Level	Ethnicity	Cohen's d	Hedges's g	Confidence Interval for Hedges's g	
Elem	Total	-0.090068	-0.090038	-0.1728409	-0.0072356
Elem	1	-0.001798	-0.001797	-0.1161588	0.1125645
Elem	2	0.201518	0.200488	-0.1458026	0.5467783
Elem	3	-0.398227	-0.397839	-0.5535445	-0.2421340
Elem	4	0.053028	0.052212	-0.4977431	0.6021675
Elem	5	0.300688	0.297454	-0.1735874	0.7684963
Middle	Total	0.183273	0.183203	0.0949444	0.2714624
Middle	1	0.427324	0.427063	0.3131124	0.5410131
Middle	2	-0.123144	-0.122310	-0.5246412	0.2800212
Middle	3	-0.223774	-0.223453	-0.4114687	-0.0354374
Middle	4	-0.044590	-0.043600	-0.7237725	0.6365734
Middle	5	0.241118	0.238762	-0.2081647	0.6856881
High	Total	0.046646	0.046628	-0.0433192	0.1365744
High	1	0.091286	0.091243	-0.0076689	0.1901550
High	2	*	*	*	*
High	3	-0.429815	-0.428080	-1.0144527	0.1582927
High	4	-0.136405	-0.131908	-0.9329098	0.6690934
High	5	-0.432429	-0.422214	-1.1327226	0.2882945

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White. * N < 9; effect sizes not calculated.

APPENDIX B

Data Used in Scatter Plot for Ethnic Groups for Change Across Year in Proportion Passing the CRCT or HSGT and Hedges's g Effect Sizes: Mathematics Achievement

				Effect Size	
Group 1	Group 2	Level	Ethnicity	Difference in Proportions	Using Hedges's g
PDS06	PDS05	Elem.	African American	−0.007856533	−0.029393810794
PDS06	PDS05	Elem.	Asian/Pacific Islander	0.027067112	−0.098713261752
PDS06	PDS05	Elem.	Hispanic	−0.047470814	0.061304766371
PDS06	PDS05	Elem.	Multiracial	−0.131830239	0.002474411014
PDS06	PDS05	Elem.	White	0.049441786	−0.042362140815
PDS06	PDS05	Middle	African American	−0.084780900	0.175348293103
PDS06	PDS05	Middle	Asian/Pacific Islander	−0.104313174	0.125796207167
PDS06	PDS05	Middle	Hispanic	−0.052178142	0.128942719575
PDS06	PDS05	Middle	Multiracial	0.078492936	0.153601127777
PDS06	PDS05	Middle	White	−0.087316176	−0.017108003355
PDS06	PDS05	High	African American	−0.021873857	0.119847832341
PDS06	PDS05	High	Asian/Pacific Islander	0.035134744	0.303735957452
PDS06	PDS05	High	Hispanic	−0.070187886	0.303141649517
PDS06	PDS05	High	Multiracial	−0.100000000	0.040951547754
PDS06	PDS05	High	White	−0.103448276	0.591999492350
CS 06	CS 05	Elem.	African American	−0.026761995	−0.072191729465
CS 06	CS 05	Elem.	Asian/Pacific Islander	−0.026101444	−0.025931299258
CS 06	CS 05	Elem.	Hispanic	−0.010341690	−0.073637099092
CS 06	CS 05	Elem.	Multiracial	0.053981107	0.008854299547
CS 06	CS 05	Elem.	White	−0.005245902	0.106186491278
CS 06	CS 05	Middle	African American	−0.055008210	−0.210674584387
CS 06	CS 05	Middle	Asian/Pacific Islander	−0.075350140	−0.114950642874
CS 06	CS 05	Middle	Hispanic	−0.084998434	−0.152546016340
CS 06	CS 05	Middle	Multiracial	−0.147086031	−0.346624789424
CS 06	CS 05	Middle	White	−0.122141119	−0.030447476868
CS 06	CS 05	High	African American	−0.065879624	−0.243781106120
CS 06	CS 06	High	Asian/Pacific Islander	−0.047898913	−0.368053157142
CS 06	CS 06	High	Hispanic	−0.035393361	−0.166294652869
CS 06*	CS 06	High	Multiracial	−0.887362637	−0.485051043214
CS 06	CS 06	High	White	0.002607762	−0.414470099010

Note. * These pairs of scores were considered to be outliers and were deleted from the data set for the primary analysis. Correlations are reported with and without the outliers in the data set.

Data Used in Scatter Plot for Ethnic Groups for Change Across Year in Proportion Passing the CRCT or HSGT and Hedges's g Effect Sizes: Science Achievement

				Effect Size	
Group 1	Group 2	Level	Ethnicity	Difference in Proportions	Using Hedges's g
PDS06	PDS05	Elem.	African American	−0.067077027	0.209344910460
PDS06	PDS05	Elem.	Asian/Pacific Islander	−0.050008994	0.070649911350
PDS06	PDS05	Elem.	Hispanic	−0.008137065	0.056428747214
PDS06	PDS05	Elem.	Multiracial	−0.168845316	0.141181441652
PDS06	PDS05	Elem.	White	−0.004761905	−0.014189133492
PDS06	PDS05	Middle	African American	0.010994265	0.017622267881
PDS06	PDS05	Middle	Asian/Pacific Islander	0.015799344	0.073131724782
PDS06	PDS05	Middle	Hispanic	−0.011243892	−0.038220199321
PDS06	PDS05	Middle	Multiracial	0.062794349	0.050407796617
PDS06	PDS05	Middle	White	−0.087775735	0.188137189809
PDS06	PDS05	High	African American	0.003424675	0.095363956493
PDS06	PDS05	High	Asian/Pacific Islander	0.082879801	−0.209810200383
PDS06	PDS05	High	Hispanic	0.028244877	0.070636316388
PDS06*	PDS05	High	Multiracial	0.188235294	−0.303959337591
PDS06	PDS05	High	White	−0.054545455	0.208039700235
CS 06	CS 05	Elem.	African American	−0.079439941	−0.166220874217
CS 06	CS 05	Elem.	Asian/Pacific Islander	−0.021276596	−0.170895891155
CS 06	CS 05	Elem.	Hispanic	−0.069731921	−0.263973841761
CS 06	CS 05	Elem.	Multiracial	0.079545455	−0.027447368835
CS 06	CS 05	Elem.	White	−0.022556391	−0.044359103696
CS 06	CS 05	Middle	African American	0.050353240	0.152244653290
CS 06*	CS 05	Middle	Asian/Pacific Islander	−0.349019608	−0.202333205571
CS 06	CS 05	Middle	Hispanic	0.009896649	0.062856144110
CS 06	CS 05	Middle	Multiracial	0.002775208	−0.089659239597
CS 06	CS 05	Middle	White	−0.049148418	−0.108850845157
CS 06	CS 05	High	African American	0.099290516	0.179591334272
CS 06*	CS 05	High	Asian/Pacific Islander	0.586580087	1.018261574088
CS 06	CS 05	High	Hispanic	−0.090476190	−0.221967206448
CS 06	CS 05	High	Multiracial	−0.007692308	0.200956643277
CS 06	CS 05	High	White	0.001788269	0.081082813372

Note. * These pairs of scores were considered to be outliers and were deleted from the data set for the primary analysis. Correlations are reported with and without the outliers in the data set.

APPENDIX C

Proportions of Students' Passing by Ethnic Group for Baseline and Year 1: Mathematics

				Proportion Passing		
Group 1	Group 2	Level	Ethnicity	Group1	Group2	Cohen's h
PDS06	PDS05	Elem.	1	0.79467	0.80253	-0.02
PDS06	PDS05	Elem.	2	0.94086	0.91379	0.10
PDS06	PDS05	Elem.	3	0.76892	0.81639	-0.12
PDS06	PDS05	Elem.	4	0.77586	0.90769	-0.37
PDS06	PDS05	Elem.	5	0.90909	0.85965	0.16
PDS06	PDS05	Middle	1	0.68157	0.76636	-0.19
PDS06	PDS05	Middle	2	0.87037	0.97468	-0.42
PDS06	PDS05	Middle	3	0.54242	0.59459	-0.11
PDS06	PDS05	Middle	4	0.69388	0.61538	0.17
PDS06	PDS05	Middle	5	0.76563	0.85294	-0.22
PDS06	PDS05	High	1	0.77893	0.80081	-0.05
PDS06	PDS05	High	2	0.97479	0.93966	0.18
PDS06	PDS05	High	3	0.73446	0.80465	-0.17
PDS06	PDS05	High	4	0.80000	0.90000	-0.28
PDS06	PDS05	High	5	0.89655	1.00000	-0.65
CS 06	CS 05	Elem.	1	0.79915	0.82591	-0.07
CS 06	CS 05	Elem.	2	0.90541	0.93151	-0.10
CS 06	CS 05	Elem.	3	0.87909	0.88943	-0.03
CS 06	CS 05	Elem.	4	0.94872	0.89474	0.20
CS 06	CS 05	Elem.	5	0.88000	0.88525	-0.02
CS 06	CS 05	Middle	1	0.57143	0.62644	-0.11
CS 06	CS 05	Middle	2	0.89524	0.97059	-0.31
CS 06	CS 05	Middle	3	0.65049	0.73548	-0.18
CS 06	CS 05	Middle	4	0.76596	0.91304	-0.41
CS 06	CS 05	Middle	5	0.74453	0.86667	-0.31
CS 06	CS 05	High	1	0.75533	0.82121	-0.16
CS 06	CS 05	High	2	0.92771	0.97561	-0.23
CS 06	CS 05	High	3	0.82927	0.86466	-0.10
CS 06	CS 05	High	4	0.07692	0.96429	-2.20
CS 06	CS 05	High	5	0.98374	0.98113	0.02

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White. Group 1 = Year 1. Group 2 = baseline. Rows in this table correspond respectively to rows in the tables in Appendix B.

Proportions of Students' Passing by Ethnic Group for Baseline and Year 1: Science

Group 1	Group 2	Level	Ethnicity	Proportion Passing Group1	Proportion Passing Group2	Cohen's h
PDS06	PDS05	Elem.	1	0.72177	0.78884	-0.16
PDS06	PDS05	Elem.	2	0.87156	0.92157	-0.17
PDS06	PDS05	Elem.	3	0.70825	0.71639	-0.02
PDS06	PDS05	Elem.	4	0.79412	0.96296	-0.55
PDS06	PDS05	Elem.	5	0.92857	0.93333	-0.02
PDS06	PDS05	Middle	1	0.75492	0.74393	0.03
PDS06	PDS05	Middle	2	0.82593	0.81013	0.04
PDS06	PDS05	Middle	3	0.56984	0.58108	-0.02
PDS06	PDS05	Middle	4	0.75510	0.69231	0.14
PDS06	PDS05	Middle	5	0.88281	0.97059	-0.35
PDS06	PDS05	High	1	0.42255	0.41913	0.01
PDS06	PDS05	High	2	0.64220	0.55932	0.17
PDS06	PDS05	High	3	0.32370	0.29545	0.06
PDS06	PDS05	High	4	0.58824	0.40000	0.38
PDS06	PDS05	High	5	0.76364	0.81818	-0.13
CS 06	CS 05	Elem.	1	0.70870	0.78814	-0.18
CS 06	CS 05	Elem.	2	0.89362	0.91489	-0.07
CS 06	CS 05	Elem.	3	0.83857	0.90830	-0.21
CS 06	CS 05	Elem.	4	0.95455	0.87500	0.29
CS 06	CS 05	Elem.	5	0.92982	0.95238	-0.10
CS 06	CS 05	Middle	1	0.66961	0.61925	0.11
CS 06	CS 05	Middle	2	0.53333	0.88235	-0.80
CS 06	CS 05	Middle	3	0.66796	0.65806	0.02
CS 06	CS 05	Middle	4	0.87234	0.86957	0.01
CS 06	CS 05	Middle	5	0.81752	0.86667	-0.14
CS 06	CS 05	High	1	0.52364	0.42435	0.20
CS 06	CS 05	High	2	0.75325	0.16667	1.26
CS 06	CS 05	High	3	0.40952	0.50000	-0.18
CS 06	CS 05	High	4	0.69231	0.70000	-0.02
CS 06	CS 05	High	5	0.91845	0.91667	0.01

Note. Ethnicity codes: 1 = African American; 2 = Asian/Pacific Islander; 3 = Hispanic; 4 = Multiracial; 5 = White. Group 1 = Year 1. Group 2 = baseline. Rows in this table correspond respectively to rows in the tables in Appendix B.

APPENDIX D

Means Table for CRCT

		Mathematics Scale Scores CRCT					
		Professional Development Schools			Comparison Schools		
		YR 2005	YR 2006	Difference	YR 2005	YR 2006	Difference
Elementary	*M*	324.90	325.13	0.23	329.12	330.82	1.70
	SD	30.824	29.031		32.882	30.771	
	N	2279	2180		1760	1814	
Middle	*M*	312.70	316.37	3.67	309.36	312.53	3.17
	SD	31.647	29.128		29.198	29.312	
	N	2732	1031		2505	953	
High	*M*	517.83	522.19	4.36	522.75	528.56	5.81
	SD	22.661	26.713		34.162	29.19	
	N	1143	1145		1261	1445	

		Science Scale Scores CRCT					
		Professional Development Schools			Comparison Schools		
		YR 2005	YR 2006	Difference	YR 2005	YR 2006	Difference
Elementary	*M*	314.51	317.22	2.71	315.18	319.24	4.06
	SD	23.548	22.506		26.089	22.337	
	N	1303	1204		1039	1050	
Middle	*M*	311.85	312.31	0.46	310.39	308.19	-2.20
	SD	22.342	23.61		20.332	21.19	
	N	2732	1031		2505	953	
High	*M*	497.12	497.56	0.44	505.4	496.65	-8.75
	SD	24.964	19.527		23.818	19.489	
	N	1161	983		1253	919	