


2007

An Investigation of the Impact of the Success for all Whole-School Reform Model on the Elementary School Proficiency Assessment and the New Jersey Assessment of Skills and Knowledge in an Urban District

Winston D. Jackson
Seton Hall University

Follow this and additional works at: <http://scholarship.shu.edu/dissertations>

 Part of the [Curriculum and Instruction Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Educational Psychology Commons](#)

Recommended Citation

Jackson, Winston D., "An Investigation of the Impact of the Success for all Whole-School Reform Model on the Elementary School Proficiency Assessment and the New Jersey Assessment of Skills and Knowledge in an Urban District" (2007). *Seton Hall University Dissertations and Theses (ETDs)*. 231.

<http://scholarship.shu.edu/dissertations/231>

AN INVESTIGATION OF THE IMPACT OF THE SUCCESS FOR ALL WHOLE-
SCHOOL REFORM MODEL ON THE ELEMENTARY SCHOOL PROFICIENCY
ASSESSMENT AND THE NEW JERSEY ASSESSMENT OF SKILLS AND
KNOWLEDGE IN AN URBAN DISTRICT

By

Winston D. Jackson

Dissertation Committee

Anthony J. Colella, Ph.D., Mentor

Karen E. Boroff, Ph.D.

James M. Caulfield, Ed.D.

Soundaram Ramaswami, Ph.D.

Submitted in partial fulfillment of the
Requirement for Degree of Doctor of Education
Seton Hall University
2006

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

APPROVAL FOR SUCCESSFUL DEFENSE

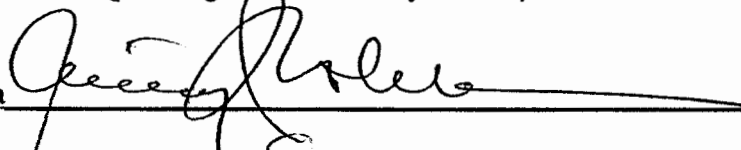
Doctoral Candidate, **Winston Jackson**, has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Fall Semester 2006**.

DISSERTATION COMMITTEE

(please sign and date beside your name)

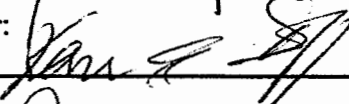
Mentor:

Dr. Anthony Colella



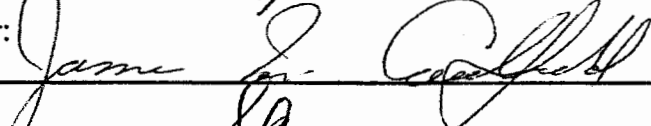
Committee Member:

Dr. Karen Boroff




Committee Member:

Dr. James Caulfield



Committee Member:

Dr. Soundaram Ramaswami



External Reader:

Associate Dean:

The mentor and any other committee members, who wish a review to recommend revision, will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file.

ACKNOWLEDGEMENTS

Thank you to my dissertation committee, Dr. Anthony J. Colella, Dr. James M. Caulfield, Dr. Soundaram Ramaswami and Dr. Karen E. Boroff, for mentoring and assisting me in completing this dissertation. Thanks to Evans Onsongo for assisting me with my statistics. I give additional thanks to Dr. Ramaswami for her valuable assistance.

I thank the following people for having a positive impact on my personal life and my professional career: Valerie, Ronald, Andrea and Keith, my brothers and sisters, because the experiences that we shared as children had the most significant impact on creating the person that I am today. John Grant, Claude Sherman and Darryl White, for being positive male role models for me. Dr. Karen E. Boroff, Ronald Callisi, Esq. and Dr. James M. Caulfield for providing me with the opportunities and the support to achieve my educational goals. Anne Boyd-Hinton, A.J. Reinoso, Ivan L. Holmes, Dr. Antonio Lewis, Dr. Glenda Johnson-Green, Anzella K. Nelms and Marion A. Bolden for providing me with invaluable administrative opportunities and experiences. Jason Cox, Sandy Sykes, Jarteau Israel, Stan Barber and Gunther Pergher for being my best friends and for supporting my efforts and achievements. Dr. Willie K. Hill, my great aunt, for leading the way with her academic achievements. Finally, Eileen Jackson, my mother, for her unconditional love and for believing that I could accomplish anything that I wanted to do.

DEDICATION

This dissertation is dedicated to my wife, Andrea C. Jackson, the most important person in my life. Thank you for allowing me put my studies ahead of you for a few years. I sincerely appreciate the sacrifices that you have made in order to allow us to accomplish this monumental undertaking. Thanks for your understanding, support and unconditional love. I love you.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	i
DEDICATION.....	ii
LIST OF TABLES.....	iii
ABSTRACT	iiii
I. INTRODUCTION.....	1
Statement of the Problem.....	9
Significance of the Study.....	10
Purpose of the Study.....	12
Research Questions.....	13
Definitions of Terms.....	14
Limitations.....	15
Summary.....	16
II. LITERATURE REVIEW.....	17
Introduction.....	17
Early Education Reform.....	18
School-wide Reform Approaches.....	23
School Development Program (SDP)/Comer.....	26
Coalition of Essential Schools.....	28
Success for All.....	32
School Reform Support and Funding.....	39
Baltimore City Public Schools Study.....	41
Charleston County Schools Study.....	43
New American Schools Study.....	46
Singer Study	49
Current and Future Whole-School Reform.....	51
Conceptual Framework.....	52
Summary.....	54
III. METHODOLOGY.....	57
Introduction.....	57
ESPA/NJ ASK4.....	57
Description of Sample/Population.....	58
Data Collection.....	58
Data Analysis.....	59
Summary.....	62
IV. DATA ANALYSIS.....	63
Introduction.....	63
Research Question 1.....	65

Research Question 2.....	67
Research Question 3.....	69
Research Question 4.....	72
Research Question 5.....	74
Research Question 6.....	78
Research Question 7.....	83
Research Question 8.....	85
Research Question 9.....	88
Research Question 10.....	94
Research Question 11.....	101
Research Question 12.....	105
V. SUMMARY, CONCLUSIONS, RECOMMENDATIONS.....	111
Summary.....	111
Conclusions.....	117
Policy Recommendations.....	119
Future Research.....	120
REFERENCES.....	122
APPENDICES	
Appendix A List of 30 Abbott Districts.....	126
Appendix B Urban Public Schools ESPA/NJ ASK4, GEPA, and HSPA Assessment Data.....	127
Appendix C New Jersey State/NCLB District Level Benchmarks/AYP.....	129
Appendix D ESPA LAL Means for WSR Models 1999-2001.....	135
Appendix E Chart of Literature Review	136

LIST OF TABLES

1. Past School Reform Approaches and Theoretical Assumptions.....	20
2. Percentage of Charleston County School District SFA Schools Students Achieving the Basic Skills Assessment Program District Standards from 1989- 1992.....	45
3. Conceptual Framework For This Study.....	52
4. T-test For District Language Arts Literacy Means For 2002-2003.....	65
5. T-test For District Language Arts Literacy Means For 2003-2004.....	65
6. T-test For District Language Arts Literacy Means For 2002-2004.....	66
7. T-test For District Reading Cluster Means For 2002-2003.....	67
8. T-test For District Reading Cluster Means For 2003-2004.....	68
9. T-test For District Reading Cluster Means For 2002-2004.....	68
10. T-tests For School Groups Language Arts Literacy for 2002.....	70
11. T-tests For School Groups Language Arts Literacy for 2003.....	70
12. T-tests For School Groups Language Arts Literacy for 2004.....	71
13. T-tests For School Groups Reading Cluster for 2002.....	72
14. T-tests For School Groups Reading Cluster for 2003.....	73
15. T-tests For School Groups Reading Cluster for 2004.....	73
16. T-test For Non-SFA Language Arts Literacy Means For 2002-2003.....	74
17. T-test For Non-SFA Language Arts Literacy Means For 2003-2004.....	75
18. T-test For Non-SFA Language Arts Literacy Means For 2002-2004.....	75
19. T-test For SFA Schools Language Arts Literacy Means For 2002-2003.....	76
20. T-test For SFA Schools Language Arts Literacy Means For 2003-2004.....	76
21. T-test For SFA Schools Language Arts Literacy Means For 2002-2004.....	76
22. T-test For Non-SFA Schools Reading Cluster Means For 2002-2003.....	79
23. T-test For Non-SFA Schools Reading Cluster Means For 2003-2004.....	79
24. T-test For Non-SFA Schools Reading Cluster Means For 2002-2004.....	79
25. T-test For SFA Schools Reading Cluster Means For 2002-2003.....	80
26. T-test For SFA Schools Reading Cluster Means For 2003-2004.....	80
27. T-test For SFA Schools Reading Cluster Means For 2002-2004.....	81
28. Non-SFA and SFA Increases and Decreases for Language Arts Literacy and Reading.....	83
29. T-tests For Non-SFA and SFA Schools Language Arts Literacy For 2002.....	84
30. T-tests For Non-SFA and SFA Schools Language Arts Literacy For 2003.....	84
31. T-tests For Non-SFA and SFA Schools Language Arts Literacy For 2004.....	85
32. T-tests for Non-SFA and SFA Schools Reading Cluster for 2002.....	86
33. T-tests for Non-SFA and SFA Schools Reading Cluster for 2003.....	86
34. T-tests for Non-SFA and SFA Schools Reading Cluster for 2004.....	87
35. Regression For Language Arts Literacy For 2002.....	89
36. Regression For Language Arts Literacy For 2003.....	91
37. Regression For Language Arts Literacy For 2004.....	93
38. Regression For Reading Cluster For 2002.....	96
39. Regression For Reading Cluster For 2003.....	98
40. Regression For Reading Cluster For 2004.....	99
41. Regression For Non-SFA Reading Cluster For 2002.....	102

42. Regression For Non-SFA Reading Cluster For 2003.....	103
43. Regression For Non-SFA Reading Cluster For 2004.....	104
44. Regression For SFA Reading Cluster For 2002.....	106
45. Regression For SFA Reading Cluster For 2003.....	108
46. Regression For SFA Reading Cluster For 2004.....	109

ABSTRACT

This study evaluates the impact of the SFA (Success for All) whole-school reform model on student achievement in an urban school district from 2002 to 2004. The ESPA (Elementary School Proficiency Assessment) and NJ ASK4 (New Jersey Assessment of Skills and Knowledge) Language Arts Literacy and Reading Cluster mean scores of 4th grade students were analyzed during this study. The mean scores were analyzed using t-tests and multiple regressions. The findings included significant Language Arts Literacy and Reading Cluster mean increases from 2002 to 2004 for SFA and Non-SFA schools. The findings also included significant Language Arts Literacy and Reading Cluster mean differences between SFA schools and Non-SFA schools.

CHAPTER I

Introduction

The Commissioner proposed that elementary schools in the Abbott districts undergo “whole-school reform,” a comprehensive approach to education that fundamentally alters the way in which decisions about education are made. A school implements whole-school reform by integrating reform throughout the school as a total institution, rather than simply adding reforms piecemeal. If carried out successfully, whole-school reform affects the culture of the entire school, including instruction, curriculum, and assessment. The reform covers education from the earliest levels, including pre-school, and can be particularly effective in enabling the disadvantaged children in poor urban communities to reach higher educational standards. The Commissioner’s recommended version of whole-school reform for elementary schools is Success For All (SFA), a nationally proven program that addresses the reading deficits of low-income, at-risk public school children (“Abbott v. Burke,” 1998).

The New Jersey Supreme Court’s Abbott decision, mandating the implementation of whole-school reform (WSR), had a profound effect on public school education in the State of New Jersey.

In arriving at its decision, the Court directed the Department of Education to study all of the various approaches to improving the academic achievement of students from low-income families. Based on its extensive review of programs and research across the nation, the department proposed “whole-school reform” as being far and away the most effective approach (N. J. D. o. Education, 2004a).

Whole-School Reform attempts to improve student achievement by involving parents and communities in school management decisions, improving the delivery of instruction and improving the curriculum. “Whole-school Reform is a complete restructuring of an entire school, putting in place a series of programs and strategies that have been proven by research to be effective” (“Guide for Implementing Urban Education Reform in Abbott Districts,”; Services, 2000).

The New Jersey Supreme Court’s Abbott decision also acknowledged that several schools districts were performing poorly. Thirty school districts in New Jersey were designated Abbott Districts (see Appendix A); consequently, they were required to adopt a whole-school reform model. Abbott district status is determined primarily by its District Factor Grouping (DFG).

The Department of Education devised this system in 1974 to rank every district in the state by seven factors: 1) per capita income level; 2) occupation level; 3) education level; 4) percent of residents below the federal poverty level; 5) density (the average number of persons per household); 6) urbanization (percent of district considered urban); and 7) unemployment (percent of those in the workforce who received some unemployment compensation). The DFG model was revised following the 1990 census, as follows: 1) Percent of population with no high school diploma; 2) Percent with some college; 3) Occupation; 4) Population density; 5) Income; 6) Unemployment; and 7) Poverty. Most Abbott Districts are found in DFG A, with a few in DFG B, and two in DFG C. Not all DFG A districts are Abbotts (Librera, 2003).

The Department of Education adopted the following economic indicators to guide Abbott classification. The district must be District Factor Group A and satisfy the

following criteria. District Factor Group B districts that also meet the following criteria may be classified as Abbott if they demonstrate additional substantial economic hardship:

1. The district must have a low-income concentration (as measured by eligibility for free lunch under the United States Department of Agriculture free/reduced lunch program) of at least 40 percent;
2. If the district has a low-income concentration less than 60 percent, then it must have an equalized value per capita of at least 3 percent below the state average and an equalized tax rate at least 30 percent greater than the state average;
3. If the low-income concentration is at least 60 percent, then the equalized value per capita must be at least 3 percent below the state average; the equalized tax rate does not factor into the eligibility requirement for these districts; and the district must be included in the NJRA eligible list (Librera, 2003).

The Urban Public Schools are an Abbott district that has a District Factor Group (DFG) rating of A.

Currently, there are thirty-one Abbott districts; Salem City is the most recent school district designated an Abbott district. The Abbott Districts are provided with the additional state funding necessary to fully implement their whole-school reform models. In addition, the New Jersey Department of Education provides the Abbott Districts with administrative expertise and support to help them effectively implement their whole-school reform models. Prior to the State of New Jersey mandating whole-school reform (WSR), the New Jersey Department of Education (NJDOE) dealt with failing and

ineffective school districts by providing them with millions of dollars in additional funding, taking them over or assigning state monitors to assist and advise superintendents and administrators.

Initially, the New Jersey Department of Education indicated that only Success for All (SFA) and five other whole-school reform models could be adopted.

Particularly, the New Jersey Department of Education identified these research-based programs: Success for All (SFA), developed by Dr. Robert E. Slavin at Johns Hopkins University; Comer School Development Program, developed by Dr. James Comer at Yale University; Adaptive Learning Environments Model (ALEM), developed by Dr. Margaret Wang at Temple University; Accelerated Schools, developed by Dr. Henry Levin at Stanford University; and Modern Red Schoolhouse, developed by a collaboration of several researchers ("Abbott v. Burke," 1998, p. 35).

In addition to these six initial models, several other elementary school models were eventually approved for adoption by the New Jersey Department of Education. The other models that could be adopted were America's Choice, Ventures, Co-NECT, Community for Learning, Coalition of Essential Schools and Alternative Program Design.

Models have been characterized by the state as either content models (i.e. prescriptive models such as SFA that provide curricula and instructional strategies to staff) or process models (i.e. non-prescriptive models such as Comer which focus on governance mechanisms, operations, and guiding principles) (Muirhead, Tyler, & Hamilton, 2001, p. 14).

As previously indicated, the implementation of whole-school reform models should significantly change school culture decision-making, curriculum, delivery of

instruction and student achievement. This change should be significant whether or not a process or content model is implemented.

The implementation of whole-school reform models was preceded by the adoption of the New Jersey Core Curriculum Content Standards (NJCCCS) in 1996.

The CCCS define the concept of “thorough and efficient” as the state constitutional guarantee that students are to be educated within a thorough and efficient system of free public schools. The CCCS describe what all students should know and be able to do upon completion of a thirteen-year public education, K-12 ("Guide for Implementing Urban Education Reform in Abbott Districts,").

The Jersey Core Curriculum Content Standards focus on the following academic areas: Visual and Performing Arts, Comprehensive Health and Physical Education, Language Arts Literacy, Mathematics, Science, Social Studies and World Languages. The New Jersey Department of Education mandated the administration of three criterion-referenced tests to measure how well students were achieving regarding the New Jersey Core Curriculum Content Standards. In 1999, the New Jersey Department of Education began requiring the administration of the Elementary School Proficiency Assessment (ESPA) to assess fourth grade students, the Grade Eight Proficiency Assessment (GEPA) to assess eighth grade students and the High School Proficiency Assessment (HSPA) to assess eleventh grade students. In 2003, the New Jersey Department of Education began administering the New Jersey Assessment of Skills and Knowledge 3 (NJ ASK3) to third grade students and the New Jersey Assessment of Skills and Knowledge 4 (NJ ASK4) to fourth grade students. The NJ ASK4 replaced the ESPA as the fourth grade test.

The ESPA, NJ ASK4, GEPA and the HSPA results (see Appendix B) were also used as a means to assess the effectiveness of elementary, middle and high schools. The ESPA and NJ ASK4 are used to determine the effectiveness of elementary schools, the GEPA is used to determine the effectiveness of middle schools and HSPA is used to determine the effectiveness of high schools. The ESPA, NJ ASK4, GEPA and the HSPA results of all New Jersey schools are annually presented on the New Jersey Department of Education website and in many statewide and local publications. The Star Ledger annually presents the statewide assessment data of all public schools and charter schools. Consequently, the public dissemination of the ESPA, NJ ASK4, GEPA and the HSPA scores has resulted in the assessments becoming the most extensively used method to determine the effectiveness of elementary, middle and high schools in the State of New Jersey.

The Urban Public Schools are evaluated in this study. Since 1998, the Urban Public Schools have adopted various elementary, middle and high school whole-school reform models. The elementary school models that have been adopted are Accelerated Schools, America's Choice, Comer/School Development Program, Community For Learning and Success for All. The middle school models that have been adopted are America's Choice, Adaptive Learning Environments Model, Comer/School Development Program and Co-NECT. The high school models that have been adopted are Alternative Program Design, Talent Development, Career Academy and Coalition of Essential Schools. Many people have used the results of the ESPA, NJ ASK4, GEPA and the HSPA to evaluate the effectiveness of these whole-school reform models because the

strategy of implementing whole-school reform promised to improve student achievement and school effectiveness.

The Urban Public Schools are often criticized for being educationally ineffective and financially inefficient. The percentage of students in the Urban Public Schools achieving proficiency on the ESPA, NJ ASK4, GEPA and the HSPA have increased in some areas and decreased in others. On the Language Arts Literacy and Mathematics sections on the ESPA and the NJ ASK4, the proficiency rates have been inconsistent, increasing some years and decreasing in others. However, despite the fact that greater percentages of students are proficient today than in 1998, unacceptable percentages of students remain partially proficient in all areas.

The ESPA scores are reported as scale scores in each of the content areas. The scores range from 100-199 (Partially Proficient), 200-249 (Proficient), and 250-300 (Advanced Proficient). The scores of students who are included in the Partially Proficient level are considered to be below the state minimum of proficiency and those students may be most in need of instructional support (N. J. S. D. O. Education, 2002, p. i). Partially proficient scores are equivalent to failing scores. The existence of unacceptable percentages of students failing the ESPA, NJ ASK4, GEPA and HSPA has caused supporters and critics of the Urban Public Schools to question the implementation and the effectiveness of whole-school reform in the district.

The No Child Left Behind Act (NCLB) is currently having a significant impact on determining the effectiveness of whole-school reform in New Jersey. “The No Child Left Behind Act of 2001 (No Child Left Behind) is a landmark in education reform designed to improve student achievement and change the culture of America’s schools” (U. S. D.

o. Education, 2004). “It is built on four common-sense pillars: accountability for results; an emphasis on doing what works based on scientific research; expanded parental options; and expanded local control and flexibility” (U. S. D. o. Education, 2004). The No Child Left Behind legislation... “called for more federal funding for Title I, as well as for more strict state and local accountability for improving education-including penalties for failing schools and new options for students attending them” (Gordon, 2003, p. 127). The most important mandate of No Child Left Behind is that 100% of the students must be proficient in Language Arts Literacy and Mathematics by the end of the 2014 school year.

Abbott schools and non-Abbott schools that fail to make adequate yearly progress (AYP) are considered low performing schools. The adequate yearly progress benchmarks (see Appendix C) are mandated by the No Child Left Behind Act and they are established by the New Jersey Department of Education. Adequate yearly progress is determined by achieving the state benchmarks or safe harbor on the ESPA, NJ ASK4, GEPA and HSPA. Safe harbor is computed by achieving an increase of 10% (or greater) of the percentage of students who failed (partially proficient) the test. Despite whole-school reform, many of the Abbott districts and Abbott schools have failed to make adequate yearly progress.

The failure to meet adequate yearly progress by schools using WSR models has resulted in the effectiveness of whole-school reform models as a strategy to improve failing schools being questioned in school districts throughout the United States. Therefore, several studies have been conducted throughout the United States regarding the effectiveness of whole-school reform models. The SFA whole-school reform model is probably the most widely used whole-school reform model in the United States and

probably the most widely criticized also. The effectiveness of the SFA model has been validated by some researchers and invalidated by others. The developers of SFA generally indicate that their studies show improvement regarding the reading levels of students and standardized test scores based on the level of implementation. Critics of SFA have generally indicated that their studies have shown that the implementation of SFA is not improving the reading levels or standardized test scores of students. These conflicting results regarding the effectiveness of SFA to improve student achievement and school effectiveness is the basis of this study.

Statement of the Problem

“In 1995, the New Jersey Department of Education decided that 28 public school districts ...were failing and that the students were not receiving a ‘thorough and efficient education’” (Public School Education Act of 1975). In 1997, the Abbott Districts, including the Urban Public Schools “...were mandated to undergo Whole-School Reform in order to provide students with a ‘thorough and efficient education’” (Public School Education Act of 1975). From 1995 to the present, the perception is that the Urban Public Schools are continuing to fail and that the students are not achieving academically. The research-based, whole-school reform models that Abbott Districts were mandated to adopt were expected to improve student achievement in the failing schools. Eighteen elementary schools in the Urban Public School have adopted SFA as their whole-school reform model. Despite their adoption of SFA, some of these schools continue to be viewed as failures by the New Jersey Department of Education and the public.

During the last few years, the Urban Public Schools also began to question whether SFA could improve test scores. SFA schools regularly failed to achieve adequate

yearly progress and their ESPA/NJ ASK4 scores were the lowest in the district. As a result of the SFA schools failure to consistently meet adequate yearly progress, the New Jersey Department of Education allowed several school districts, including the Urban Public Schools, to discontinue the implementation of the SFA whole-school reform model after the 2003-2004 school year. The New Jersey Department of Education has allowed the Urban Public Schools to implement an Alternative WSR model.

Significance of the Study

Success for All studies have been conducted in the Baltimore County Public Schools (BCPS) in Maryland, Baltimore City Public Schools in Maryland, Charleston County Schools (BCPS) in South Carolina, Dade County in Florida, Putnam County in Florida and other districts throughout the United States. The research suggests that some schools and school districts that have implemented SFA have improved their reading levels and their standardized tests scores. However, the research does not suggest that all SFA schools and districts have improved reading levels and standardized tests scores. In addition to the Urban Public Schools, some schools and school districts have dropped SFA because student achievement did not improve. Pogrow made the following statement regarding the Baltimore City Public Schools in Maryland.

In 1996, BCPS had the SFA reform model in place in 21 schools. Concerned about the poor results, in 1996 the district developed its own reading intervention that combined in-house materials with a commercial product. However, the SFA schools were allowed to retain SFA. The number of schools that subsequently chose to stay with SFA dropped to 11, and these schools are limiting the use of SFA to a few grade levels and have had to modify the program to meet their needs (Pogrow, 2000, p. 16).

The Urban Public Schools have made noteworthy progress regarding administrative and fiscal operations. There are some schools in the Urban Public Schools

that consistently meet and exceed NCLB adequate yearly progress benchmarks on the ESPA, NJ ASK4, GEPA or the HSPA; however, an unacceptable number of schools consistently fail to meet the benchmarks. Despite whole-school reform, the Urban Public School district did not achieve the mandatory 2002 and 2003 adequate yearly progress benchmarks in any subject on the ESPA, NJ ASK4, GEPA or the HSPA. In March of 2005, 30 school districts, including the Urban Public Schools, were identified by the New Jersey Department of Education as a NCLB-Designated District In Need of Improvement.

These districts were identified because they met three statistical criteria:

- First, when the scores of all students in the three tested grades in all schools in the district were aggregated, the districts did not make Adequate Yearly Progress (AYP) for two consecutive years (2002-2003 and 2003-2004) in the same subject at all grade levels.
- Second, when the students' scores were disaggregated by school, half the schools in the district did not make AYP in 2004.
- And third, when any of the schools that did not make AYP in 2004 achieved fewer than 90 percent of the 40 indicators each school is required to meet (N. J. D. o. Education, 2005).

There are five primary reasons why this study is significant. Firstly, this study will add to the existing knowledge base regarding the efficacy of the Success for All whole-school reform model and whole-school reform models in general. Secondly, this study will build on the existing research that suggests that there is a positive relationship between the implementation of the SFA whole-school reform and student achievement.

Thirdly, this study will provide research-based information to assist the New Jersey Department of Education in policy decisions regarding SFA as an effective whole-school reform model. Fourthly, this study will provide the Urban Public Schools with research regarding the impact of the SFA whole-school reform model on student achievement in the Urban Public Schools specifically, Language Arts Literacy and Reading on the ESPA and the NJ ASK4. Fifthly, this study will provide information that will help the Urban Public Schools evaluate the decision to cease implementing the SFA whole-school reform after the 2003-2004 school year.

Purpose of the Study

The purpose of this study is to investigate the impact of the SFA whole-school reform model on student achievement as measured by the ESPA and NJ ASK4 in the Urban Public Schools. The study will result in findings that will help to determine the impact of the SFA whole-school reform model on Language Arts Literacy mean scores and Reading Cluster mean scores. In addition, the study will also establish the effectiveness of SFA and attempt to determine its overall impact on student achievement in the district from 2002-2004 across school groups. This study seeks to determine if student achievement has improved overall among the cohorts district wide from 2002-2004. Also, this study seeks to determine if student achievement differs among school groups where one school group has schools with only SFA and other school groups have predominately other models. This study seeks to determine if student achievement differs between SFA and non-SFA schools. This study also seeks to determine the impact of several variables on student achievement such as gender, ethnicity and WSR model.

Research Questions

The following research questions will be answered during this study:

1. Are there significant differences among the cohorts when comparing Language Arts Literacy mean scores from 2002-2004?
2. Are there significant differences among the cohorts when comparing Reading Cluster mean scores from 2002-2004?
3. Are there significant differences between School Group A and other School Groups when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
4. Are there significant differences between School Group A and other School Groups when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?
5. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Language Arts Literacy mean scores from 2002-2003, 2003-2004 and 2002-2004?
6. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Reading Cluster mean scores from 2002-2003, 2003-2004 and 2002-2004?
7. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?

8. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?
9. Are gender, ethnicity and the WSR model good predictors of Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
10. Are gender, ethnicity and the WSR model good predictors of Reading Cluster mean scores for school years 2002, 2003 and 2004?
11. Are gender and ethnicity good predictors of Non-SFA schools Reading Cluster mean scores for school years 2002, 2003 and 2004?
12. Are gender and ethnicity good predictors of SFA schools Reading Cluster mean scores for school years 2002, 2003 and 2004?

Definition of Terms

Cohort – Represents a group of schools from the same school year.

ESPA Elementary School Proficiency Assessment - This assessment was given to 4th Grade students in New Jersey from the 1998-1999 school year through the 2001-2002 school year.

Proficiency Rates – The analysis of the three proficiency ratings on the ESPA and the NJ ASK. The rates are partially proficient (100-199), proficient (200-249) and advanced proficient (250-300).

NJ ASK4 – New Jersey Assessment of Skills and Knowledge 4. This assessment is given to 4th Grade students in New Jersey beginning in the 2002-2003 school year.

UPS - Urban Public Schools is an urban school district located in New Jersey.

School Groups - The Urban School District is divided into five groups of schools called School Groups. School Groups A, B, C, and D consists of elementary schools and School Group E consists of secondary schools.

WSR – Whole-School Reform. A comprehensive strategy to improve an entire school, including but not limited to, school culture, curriculum, instruction, student achievement and leadership.

WSR Cohort – Represents a group of schools that implemented their whole-school reform models in the same school year.

WSR Cohort 1 - Represents a group of schools that implemented their whole-school reform models at the beginning of the 1998-1999 school year.

WSR Cohort 2 - Represents a group of schools that implemented their whole-school reform models at the beginning of the 1999-2000 school year.

WSR Cohort 2 Midyear - Represents a group of schools that implemented their whole-school reform models during middle of the 1999-2000 school year.

WSR Cohort 3 - Represents a group of schools that implemented their whole-school reform models at the beginning of the 2000-2001 school year.

WSR Cohort 3 Midyear - Represents a group of schools that implemented their whole-school reform models during middle of the 2000-2001 school year.

Limitations

This study does not disaggregate the ESPA and the NJ ASK4 student data of students who did not attend SFA schools in the Urban Public Schools for the entire time period covered by this study. Also, this study does not control for the degree of teacher or school fidelity, proficiency or experience regarding the implementation of the SFA

whole-school reform model. Teaching experience or teaching competence is not controlled in this study. In addition, this study does not include student data derived from Limited English Proficient (LEP) or Special Education students from the Urban Public Schools.

Summary

This chapter provided an overview of the conditions that mandated whole-school reform and, ultimately, the implementation of the SFA whole-school reform model in many schools throughout New Jersey. The statement of the problem section indicated that the Urban Public Schools were performing poorly despite the implementation of the SFA whole-school reform model in some of its schools. The significance section indicated that previous SFA studies in Maryland, South Carolina and Florida have demonstrated that SFA has produced mixed results regarding student achievement. This section also indicated the Urban Public Schools failed to make adequate yearly progress on the ESPA and the NJ ASK4. The purpose section indicated that this study was conducted to determine the impact of the SFA whole-school reform model on student achievement in SFA schools as measured by ESPA and NJ ASK4. This chapter also included the research questions, definition of terms and limitations of the study.

CHAPTER II

LITERATURE REVIEW

Introduction

This chapter reviews and analyzes literature that is relevant to this study. The research reviewed provides the conceptual framework for the study and analyses of empirical studies. Sections 1, 2 and 4 provide the conceptual framework for the study and sections 5, 6, 7 and 8 provide analyses of empirical studies. Section 9 provides information regarding the future of WSR and section 10 provides a conceptual framework for this study.

The first section provides an overview of early school reform strategies that are later combined to establish whole-school reform models. The second section describes and evaluates the School Development Program whole-school reform model. This section is relevant because the School Development Program was one of the first whole-school reform models and it focused on the social development of students. The third section describes and evaluates the Coalition of Essential Schools whole-school reform model. This section is relevant because the Coalition of Essential Schools model deals primarily with processes and ideals of creating effective schools. The fourth section describes and evaluates the Success for All whole-school reform model. This section is relevant because the Success for All model is evaluated in this study.

The fifth section examines an SFA study conducted in the Baltimore City Public Schools. The title of this study is called “Success for All: Longitudinal Effects of a Restructuring Program for Inner-City Elementary Schools;” however, it will be referred to as the Baltimore City Public Schools Study in this paper. This study is relevant

because it was one of the first studies conducted to evaluate SFA and it establishes the basis for the efficacy of SFA. The sixth section examines an SFA study conducted in the Charleston County Schools. The title of this study is called “Success for Some: An Evaluation of a Success for All Program;” however, it will be referred to as the Charleston County Schools Study in this paper. This study is relevant because SFA was implemented as a strategy to improve student achievement and outcomes were measured by standardized tests. This study also provides mixed results regarding the effectiveness of the SFA whole-school reform model as it relates to student achievement. The seventh section examines a study of the New American schools. The title of this study is called “Facing the Challenges of Whole-School Reform: New After a Decade;” however, it will be referred to as the New American Schools Study in this paper. This study is relevant because the developer, like the studies in sections five and six, did not conduct it. In addition, this is one of the more recent studies conducted that evaluated some SFA schools. The eighth section examines a study conducted on the Abbott Districts in New Jersey. The title of this study is called “An Outcomes Based Study of Whole School Reform in the Abbott Districts: An In-Depth Study of Academic Progress at the Lower Elementary Level Based on the Fourth Grade ESPAS;” however, it will be referred to as the Singer Study in this paper. This study is relevant because it is current and it evaluates the implementation of whole-school reform and the impact of whole-school reform on student achievement in New Jersey.

Early Education Reform

It is important to explore early education reform because it provides the basis for today’s whole-school reform models. Early education reform incorporated some of same

strategies that are used in whole-school reform today. Trump stated, “It is true literally that, so far as schools are concerned, there is today nothing new under the sun. The problem is that most schools have not put ideas together in a comprehensive way” (Trump, 1977, p. 13). The process of combining ideas together in a comprehensive manner is a fundamental concept of whole-school reform and whole-school reform models.

The process of improving schools is a dynamic process that continues to be a struggle. The terms used for this process have also changed over time. Various terms have been used to describe the processes, methods and strategies used to improve public schools. “It has been variously labeled new education, progressive education, staff utilization, school innovation, alternative schools, excellence in education, and school reform” (Keefe, Jenkins, & Hersey, 1992, p. 15). In addition, school reform has been called system reform, restructuring, whole-school reform and many other terms. Consequently, the dynamic process of school improvement will probably lead to the creation of new terminology or the recycling of old terminology in the future.

The dynamic nature of school improvement can be understood by analyzing the implementation of major curricular and instructional strategies.

In the 1920s, the “behavioral objectives” movement gained momentum. In the 1930s, attempts to sequence curricula and to use teacher aides came to the fore. The Eight-Year Study of the 1930s represented a significant national move to revitalize and change secondary schools, but the advent of World War II dulled its impact. Staff utilization studies and subject curriculum revisions of the 1950s led to the Model Schools Project (MSP - 1969 to 1974) under the direction of J. Lloyd Trump and William Georgiades (Trump, 1977; Trump and Georgiades, 1978) (Keefe et al., 1992, p. 15).

Table 1 is chart presenting these school reform approaches and the theoretical assumptions that support these approaches. These curricular and instructional strategies and others contributed to the current knowledge base leading to whole-school reform.

Table 1.

Past School Reform Approaches and Theoretical Assumptions

School Reform Approaches	Theoretical Assumptions
Behavioral Objectives	Placing emphasis "...upon the statement of specific educational objectives in behavior terms and ...the transition from statement of philosophy to statement of the objectives of the program" (Heidenreich, 1972, p. 2).
Sequence Curriculum/Teacher Aide Use	A "...unit approach with activities and large centers of interest is desirable, provided plans are made in advance for the <i>scope</i> and <i>sequence</i> to be followed" (Gwynn & Chase, 1969, p. 241). Using aides to perform non-teaching tasks.

Past School Reform Approaches and Theoretical Assumptions

School Reform Approaches	Theoretical Assumptions
Revitalize and Change Secondary Schools	An emphasis on academic choice, academic opportunities, instructional innovation and unconventional school programs. Also, unconventional methods of evaluating students productivity and school programs.
Subject Curriculum Revisions	Revisions "...emphasizing that the whole state should take part in curriculum revision and ...the attempt to develop experimental procedures and suggestive material of a source nature" (Gwynn & Chase, 1969, p. 154).

School improvement is also influenced by politics, in addition to being influenced by curricular and instructional movements. The struggle to improve schools has become more political with the No Child Left Behind mandate requiring all students to be proficient in Language Arts Literacy and Mathematics on state assessments by the end of the 2014 school year. According to Keefe, Jenkins, & Hersey (1992), "Restructuring has been a continuous political and educational goal since the beginning of public schooling" (p. 15). "Education reform movements are fueled by crisis that typically are posed in

political hyperbole: Sputnik and the race in the late 1950s, the War on Poverty in the 1960s and early 1970s, A Nation at Risk in the 1980s” (Astuto, Clark, Read, McGree, & Fernandez, 1994, p. 15).

In 1983, the National Commission on Excellence in Education (NCEE) “released a report titled, *A Nation at Risk: The Imperative for Educational Reform*. The firestorm of public reaction was so great that Reagan found it more politically advantageous to embrace the report...” (Marsh & Willis, 1999 1995, p. 83). This report was the latest example of an alleged crisis in school reform. “Education reform has marked America’s social and political landscape since the publication in 1983 of *A Nation at Risk* (National Commission on Excellence in Education)” (Rossi & Stringfield, 1997, p. iii). “*A Nation at Risk*, apparently, told the United States exactly what it was ready to hear, for public approval was instantaneous and overwhelming” (Marsh & Willis, 1999 1995, p. 83). As a result of *A Nation at Risk*, “Presidents, congressional leaders, state and local policymakers, educators, parents, and business leaders have all endorsed reforms in schools, and a wide-ranging array of reform efforts are underway currently in thousands of schools across the nation” (Rossi & Stringfield, 1997, p. iii).

According to The Carnegie Foundation (1988)

Since 1983 school reform has been at the top of the national agenda. Educators, governors, and even corporate leaders have played starring roles in this impressive drama. The curriculum was enriched, graduation standards rose, teacher salaries improved, certification procedures tightened, and pre-school education has received vigorous support (Teaching, 1988p. xi).

A Nation at Risk has created a climate that produced an increase in the implementation of whole-school reform models as a strategy to improve student achievement and schools.

The No Child Left Behind legislation of 2001 is currently creating a crisis that is fueling education reform.

School-wide Reform Approaches

Beginning in 1966, school-wide reform approaches, or whole-school reform models as comprehensive strategies to improve K-12 schools, were implemented. The earliest models were The Foxfire Fund, High/Scope and the School Development Program. The Foxfire Fund is a K-12 program that was established in 1966. “The Foxfire Fund promotes the idea of active, learner-centered, community focused education. The Foxfire “Core Practices are expected to guide instructional methods, materials, and strategies” (Herman, 1999, p. 72). High/Scope is a K-3 program that was established in 1967. “The High/Scope K-3 Model is based on the belief that children should be active participants in their own education and that they learn best from hands-on experiences. A primary goal is to improve children’s problem solving and independent thinking skills” (Herman, 1999, p. 81). The School Development Program is a K-12 program established in 1968. “The approach is based on the theory that children learn better when they form strong relationships with the adults in their lives-including parents, teachers, and members of church and other community groups-in an environment of mutual respect” (Herman, 1999, p. 110). These early school-wide improvement programs presented models for future programs.

The Model Schools Project study evaluated successful schools and combined the strategies used by these schools to recommend strategies to improve schools. “The National Association of Secondary School Principals received a substantial grant from the Danforth Foundation in May 1968...” (Trump, 1977, p. 269) to conduct the Model

Schools Project study. The Model Schools Project study was noteworthy because, “[t]he project’s purpose was to demonstrate in a group of schools that educational opportunities can be enhanced when a variety of educational innovations are assembled and coordinated over several years with improved methods of program evaluation” (Trump & Georgiades, 1978, p. 2). According to Trump (1977), “... a total of 34 schools participated in the project for varied lengths of time, most of them for six years” (p. 269). Funded by the Danforth Foundation and supported by the NASSP, the MSP advanced three reasons for the failure of previous efforts to change schools:

- Most innovations were superficial rather than real. Unless teachers and pupils and the school management team actually worked in new ways, progress would not occur.
- No school adopted all – or even most – of the proposed innovations in a systematic, total program. Potential gains in one area were nullified by conventional practices in others.
- Schools failed to evaluate all aspects of their programs in the light of competency goals and objectives (Keefe et al., 1992, p. 15).

These reasons and many more contributed to the failure of school reform.

As a result of the Model Schools Project study, the qualities of a successful school became evident. The qualities of successful schools in the Model School Project were developed into recommendations for school improvement and reform.

The Model Schools Project advocated five basic changes:

1. The principal must be the instructional leader of the school and should devote about three-fourths of his or her working time to the improvement

of instruction. (Reallocation of responsibilities with a school supervisory-management team would make it possible.)

2. The instructional staff must be reorganized. Teachers should have the help of aides who, under teacher supervision, would assume certain tasks that consume more than one-third of the teaching day.
3. Teachers need freedom for educational planning. No substantial improvement in learning can occur unless students themselves have more time for directed independent study; i.e., the opportunity to experience a variety of learning activities away from the constant supervision of teachers.
4. The curriculum must offer continuous contact with essential materials in all the basic areas of human knowledge. It should consist of materials for the real world that the student knows, as well as those from the more specialized world of the teacher.

Even if all the preceding changes were made, none would be particularly effective without better utilization of the “things”; e.g., buildings, equipment and supplies, and money (Keefe et al., 1992, p. 16).

School Development Program (SDP)/Comer Model

The School Development Program/Comer whole-school reform model is important to this study because it was one of the first whole-school reform models that was implemented and its primary focus is on the social development of students. This model also provides results that indicate that whole-school reform can be an effective strategy to improve student achievement.

James Comer, a child psychiatrist at Yale University, founded the School Development Program (SDP) in 1968. The School Development Program is often referred to as the Comer Model. “In collaboration with the New Haven School System in Connecticut, the program was first implemented in the two lowest-achieving elementary schools in New Haven in 1968” (Protheroe, 1998, p. 76). According to Protheroe (1998), “Today over 700 schools are using the program; 541 of these are elementary schools, and the rest are middle and high schools” (p. 79).

“The School Development Program is based on the belief that all children have the potential to succeed if their basic needs are met and they are challenged to do their best” (Protheroe, 1998, p. 76). The School Development Program also has the belief that all children can be high academic achievers. According to Herman (1999), “The approach is based on the theory that children learn better when they form strong relationships with the adults in their lives-including parents, teachers, and members of church and other community groups-in an environment of mutual respect” (p. 110). Herman (1999) also indicated that “the main goal of the program is to develop in students the personal, social and moral strengths necessary to achieve success in school” (p. 110). This is necessary because, “many of today’s children come to school with developmental gaps that impair their ability to learn”(*Overview of the School Development Program*, 2004).

“The program identifies six developmental pathways along which all children progress: speech and language, physical, social, psychological/emotional, moral, and cognitive/academic” (Protheroe, 1998, p. 76). The School Development Program developers believe that these six pathways are critical to human development. “The SDP is essentially ‘content free’ and, in principal, could be adapted to diverse local curricula”

(Herman & Stringfield, 1997, p. 9). “The program does not specify instructional methods; however, it does emphasize certain curricular components such as a focus on reading and mathematics, concentration on a few core academic courses for longer periods of time, and a limit on electives” (Protheroe, 1998, p. 76).

“The program is designed to build on existing school programs, with a focus on creating a supportive and understanding climate that will ensure that students are ready and motivated to learn” (Protheroe, 1998, p. 76). The School Development Program established the following nine essential components that must be implemented to create the climate that is necessary to improve student achievement.

Nine components—three guiding principles, three mechanisms for development, and three operations—are considered essential:

- “no fault” decision-making (principle)
- consensus decision-making (principle)
- collaboration (principle)
- parent involvement program (mechanism)
- school planning management team (mechanism)
- mental health team (mechanism)
- a comprehensive school plan (operation)
- staff development (operation), and
- assessment and modification (operation) (Herman & Stringfield, 1997, p. 8).

According to Protheroe (1998), “The School Planning and Management Team is the most important component of the program and the team’s basic function is to set the agenda and the social and academic climate of the school” (p. 77).

“Studies conducted by the SDP and by independent researchers indicate significant effects on school climate, student attendance, and student achievement” (*Overview of the School Development Program*, 2004). According to Herman (1999), “The School Development Program has promising research on student achievement effects” (p. 111). Herman evaluated twelve studies that were conducted by the program developers and independent evaluators. Three of these studies “...one conducted by the developers and two conducted by independent researchers-were considered rigorous...” (Herman, 1999, p. 111).

One research review, for example, found that schools using the School Development Program had significantly higher academic achievement than that of non-School Development Program schools in reading, mathematics, and language, measured by scores on the Iowa Test of Basic Skills and classroom grades. This finding is supported by the other two studies, which found that students in School Development Program schools tended to have higher test scores and grades, and to show greater improvement over time, compared to students who were not in such schools (Herman, 1999, p. 111).

Coalition of Essential Schools (CES)

The Coalition of Essential Schools whole-school reform model is important because it focuses on processes and ideals as a means to improve student achievement.

This model also provides results that indicate whole-school reform can be an effective strategy to improve student achievement.

TheodoreSizer, Professor of Education at Brown University, founded the Coalition of Essential Schools in 1984. Based on the findings of Sizer's *A Study of High Schools*, conducted from 1979 to 1984, and the widely read book, *Horace's Compromise*, Sizer delineated a set of principles to guide reform in high schools (Herman, 1999, p. 39). "CES started in five schools in 1984. As of 1996, CES listed over 1,000 schools as 'exploring,' 'planning,' or 'member schools' within CES" (Herman & Stringfield, 1997, p. 38).

According to the developer, schools that are exploring the Coalition of Essential Schools begin by examining the Common Principles, discussing their meaning in faculty meetings, and determining whether the philosophy is consistent with the school's goals. In the planning stage, which typically lasts at least a year, a school plans how it will implement the Common Principles, including any changes in scheduling and instructional practice that teachers decide will have to be made in order to support adoption of the Principles (Herman, 1999, p. 41). "In 1998, approximately 1,000 schools (more than half elementary schools) were engaged with Coalition of Essential Schools at some level of involvement; about 250 of those schools were Coalition members" (Herman, 1999, p. 39). The ranking of CES schools by level of involvement is something that is unique to the Coalition of Essential Schools.

"The initial focus of CES was on the 'triangle of learning' between teacher, student, and subject matter; the model has grown from classroom to school-based change as developers recognized that effective reform requires more fundamental changes in the

school” (Herman & Stringfield, 1997, p. 37). These fundamental changes include school design, structure, leadership and the establishment of community connections. The CES approach to effective reform “...stresses the importance of the local wisdom of the schools and recognizes the individual school as the fundamental unit of change...” (Protheroe, 1998, p. 37). “The Coalition of Essential Schools (CES) is not a specific model of school reform” (Herman, 1999, p. 39). “The program offers no specific model or program for schools to adopt. Each school develops its own programs based on the Coalition’s Common Principles and suits them to its particular students, faculty and community” (Protheroe, 1998, p. 38).

The key feature of the Coalition of Essential Schools (CES) is the set of “Common Principles” that guide school reform:

1. The school should focus on helping children learn to use their minds well.
2. The school’s goals should be simple: that each student master a limited number of essential skills and areas of knowledge.
3. The school’s goals should apply to all students.
4. Teaching and learning should be personalized to the maximum feasible extent.
5. The governing practical metaphor of the school should be student-as-worker, teacher-as-coach.
6. The diploma should be awarded upon demonstration of the central skills and knowledge of the school’s program.
7. The tone of the school should stress unanxious expectation, trust, and decency.

8. The principal and teachers should perceive themselves as generalists first and specialists second.
9. Teacher loads should be 80 or fewer pupils, and per-pupil cost should not exceed traditional school costs by more than 10 percent.
10. The school should demonstrate non-discriminatory and inclusive policies, practices, and pedagogies (Herman, 1999, p. 38).

The Common Principles explained previously are an abbreviated version; the standard version can be obtained from the developers of the Coalition of Essential Schools. “In the Coalition’s view, schools begin restructuring efforts with teachers, using the nine principles as guides, to design methods, models, and procedures for restructuring that accommodate varying local conditions” (Protheroe, 1998, p. 37).

“Coalition schools are difficult to evaluate—first, because each implementation is intended to be unique, and second, because the goals of the program are viewed by the developer as not being at all aligned with traditional norm-referenced tests” (Herman & Stringfield, 1997, p. 38). “A basic tenet of the Coalition of Essential Schools approach is that assessment should be made by evaluating students’ degree of understanding of a subject through their performance in authentic activities” (Protheroe, 1998, p. 37).

Authentic activities may include tasks or projects that are evaluated by direct teacher observation. A demonstration of competence and mastery is necessary to complete the authentic activities. “Students’ formal assessments, known as ‘exhibitions,’ often come at the end of the year and can take many forms, depending on a school or a class”

(Protheroe, 1998, p. 37). It is necessary for students to demonstrate mastery while completing a final exhibition to graduate.

According to Herman (1999), “The research on Coalition of Essential Schools is extensive with respect to implementation, but quite weak with respect to student outcomes. Of 30 available studies, six focused on student-level outcomes” (p. 39).

The research shows mixed effects on student achievement... Only two studies were rigorous enough to be considered... One study found that one school’s scores on the Delaware Educational Assessment Program rose from 46.0 to 50.5 over three years. The second study found that reading and mathematics scores on the Comprehensive Test of Basic Skills dropped over time (Herman, 1999, p. 39). Herman and Stringfield (1997) also indicated “evaluations of Coalition schools tend to show mixed achievement outcomes, but positive impact on students’ participation in school and changes in teaching approaches” (p. 38).

Success for All (SFA)

It is important to review the Success for All whole-school reform model because the impact of this model on student achievement is the focus of this study. Specifically, the effects of SFA on Language Arts Literacy and Reading Cluster mean scores will be explored in this study. This model also provides results that indicate whole-school reform can be an effective strategy to improve student achievement.

Robert Slavin and his colleagues at the John Hopkins University’s Center for the Social Organization of Schools in Baltimore, Maryland founded the Success for All model. “They designed Success for All in response to a challenge from the Baltimore City Public Schools to develop an approach that would address the problems of urban students, based on research about effective instructional practice” (Herman, 1999, p.

116). “Success For All was first implemented in the 1987-88 school year in one inner-city Baltimore school, Abbottston Elementary” (Madden, 1993, p. 125).

As of 2002, Success for All Foundation (SFAF) programs are being implemented in about 1,500 schools in over 500 districts in 48 states in all parts of the United States, Guam, and the Virgin Islands. Versions of the model are also used in other countries, including England, Israel, Canada, Mexico and Australia. (SFAF Schools, ¶ 1).

Success for All has a theoretical base that is different from the School Development Program and the Coalition of Essential Schools. The principal theoretical basis for the Success For All approach is the idea that learning deficits must be prevented in a comprehensive approach, emphasizing early education, improvement in instruction and curriculum, and intensive intervention at the earliest possible stage when deficiencies first begin to appear (Madden, 1993, p. 125). This theoretical base provides the foundation for the goals of the SFA whole-school reform model.

“The main goal of Success for All is to ensure success in reading” (Herman, 1999, p. 115). Specifically, the main goal of SFA is to provide students with the necessary skills, support, and intervention so that they will be on or above reading level by the end of the third grade. “Secondary goals include reducing the number of referrals to special education, reducing the number of students who are retained or ‘held back,’ increasing daily attendance, and addressing family needs” (Herman, 1999, p. 115).

The program emphasizes developmentally appropriate instructional programs in the primary grades to ensure that no child is allowed to “fall through the cracks.” The program focuses on minimizing retention and avoiding dependence on long-term special services, such as special education classes (Protheroe, 1998, p. 81). “Success for All is a

comprehensive approach to restructuring schools, especially those serving students placed at risk, to ensure that every child learns how to read” (Herman, 1999, p. 115).

“Success for All focuses on reading, writing, and language arts” (Protheroe, 1998, p. 81).

The strategy for implementing the SFA whole-school reform model is also different from the School Development Program and the Coalition of Essential Schools.

The approach to implementing SFA is explained in greater detail by the SFA

components. The approach has nine components:

- a reading curriculum designed to provide at least 90 minutes of daily instruction in classes regrouped across age lines according to reading performance;
 - continual assessment of student progress (at least once every eight weeks);
 - one-to-one reading tutors;
 - an Early Learning Program for prekindergarten and kindergarten that emphasizes language development and reading;
 - an emphasis on cooperative learning as a key teaching strategy;
 - a family support team to encourage parents’ support and involvement as well as to address problems at home;
 - a local facilitator to provide mentoring, counseling, and support to the school as needed;
 - staff support teams that assist teachers during the implementation process;
- and

- training and technical assistance provided by Success for All staff on such topics as reading assessment, classroom management, and cooperative learning (Herman, 1999, p. 115).

According to Herman and Stringfield, “One strength of the program is that the key elements—regrouping, assessment, individualization, remediation—are systemic; the structure of the program, rather than the personalities implementing it, are designed to maximize the probability of academic success” (p. 22).

“Success for All uses a highly structured curriculum focused on reading and English language arts” (Herman, 1999, p. 117). The language arts program is divided into four components: Curiosity Corner, Early Learning, Reading Roots and Reading Wings. Curiosity Corner and Early Learning are full day instructional programs. Reading Roots and Wings “...requires 90 minutes per day of reading instruction targeted to classes grouped by reading level beginning in grade one. Grouping is revised every eight weeks based on individual assessments of students’ reading skills” (Herman, 1999, p. 117). Based on the groups needed, the number of teachers teaching Roots or Wings can be changed during regrouping to maintain equal class sizes. “In grades one to three, specially trained, certified teachers work one-on-one with any students who are failing to keep up with their classmates in reading” (Protheroe, 1998, p. 81). Success for All requires tutoring for 50% of grade 1 students, 30% of grade 2 students and 20% of grade 3 students. Students are tutored for 20 minutes each day.

The Early Learning Program can be implemented in pre-kindergarten and kindergarten classes. According to Protheroe (1998), “The Early Learning program for pre-kindergarten students emphasizes language development, readiness, and positive self-

concept” (p. 81). “The Early Learning Program, for...kindergarten, focuses on developing oral language skills using developer-provided materials” (Herman, 1999, p. 117). In addition, “A full-day kindergarten program continues the emphasis on language, using children’s literature and big books, as well as oral and written composition, activities promoting the development of concepts about print, alphabet games, and math concept development” (Protheroe, 1998, p. 81).

The Reading Roots program is implemented in kindergarten classes at mid-year or the beginning of first grade. Reading Roots is a “beginning reading program in which students work with controlled-vocabulary mini-books and repeated oral reading. Reading Roots involves a blend of phonics and ‘whole language’ techniques and uses children’s literature and student text supplemented by teacher-read text...” (Herman, 1999, p. 117). In Roots, students are regrouped based on mastering a set of stories ranked from 1-48. For example, a class may consist of a group of students who mastered book number five. Students who are in grades higher than kindergarten may be placed in Reading Roots classes based on the SFA assessments. In addition, kindergarten and grade one students could be placed in Reading Wings classes based on the SFA assessments.

The Reading Wings program is implemented in grades two through six. Reading Wings, which uses a wide range of commonly available basals, anthologies, and novels. The developer requires teachers to use specific strategies in Reading Wings, in which students read stories to each other and discuss content and structure, as well as participate in activities on listening comprehension, vocabulary building, reading fluency, and writing (Herman, 1999, p. 117). In Wings, students are regrouped by reading level proficiency, expressed in half-year intervals. For example, a class may consist of students

who are on a 2.2 level. All of these students would be reading on a level consistent with Grade 2 students in the second half of the second grade. Students may be placed in reading groups that are higher or lower than their respective levels based on the SFA assessments.

The Success for All whole-school reform model has a full-time Reading Facilitator. "... [T]he facilitator coordinates eight-week assessments, assists the Family Support Team, facilitates staff support teams, plans and implements staff development, and school resources" (Protheroe, 1998, p. 82). The Reading Facilitator receives continuous training and support, from the Success for All Foundation, focused on maintaining and improving the implementation of the SFA program. The Reading Facilitator also attends Network Meetings and Leadership Academies to develop and improve educational leadership skills.

The Success for All whole-school reform model has a Family Support Team. The Family Support Team provides instructional and non-instructional support for students, families and teachers, in order to address issues that may prevent students from learning. Each school has a "family support team" to increase family involvement. The goal of these teams is to encourage parents to read to students, to involve parents in school, and to help families address any problems at home that affect a student's ability to learn in school (e.g., by providing referrals to social services) (Herman, 1999, p. 118). The Family Support Team serves as a pre-referral committee for the Child Study Team (CST). The Child Study Team evaluates students that are having difficulty learning, to determine if they are eligible for special education or related services.

“The research on Success for All is strong, based on its size and consistency of findings. Sixteen empirical studies, detailing information from about two dozen different sites, were available for review; all but one used sufficiently rigorous methodologies...” (Herman, 1999, p. 116).

The basic research design used to determine the effectiveness of SFA has been to match the SFA school with a control school that is similar in terms of poverty level, historical achievement level, ethnicity of students, and other factors (Slavin, Madden, Dolan, Wasik, Ross, Smith, & Dianda, 1996) (Weiler, 1998, p. 2). The SFA program developers used the method of matching SFA schools with non-SFA schools as their primary method during evaluations.

The results of these evaluations provide evidence that SFA is having a positive effect on student achievement. According to Herman & Stringfield (1997), “Longitudinal studies, using matched control students in matched schools, indicate that Success for All improves student achievement (Slavin, Karweit, and Wasik 1994)” (p. 19).

Compared to control schools, including schools using other approaches that employ one-on-one tutors, Success for All schools show significant improvements in students’ reading performance on standardized tests (e.g., Comprehensive Test of Basic Skills, Woodcock Reading Mastery Tests, and the Durrell Analysis of Reading Difficulty) and other measures (Herman, 1999, p. 116). Herman (1999) stated, “Not only does the research on Success for All show statistically and educationally significant improvement in student scores, but it does so consistently across the studies reviewed” (p. 116). Herman (1999) also stated, “The approach has the greatest effect on students scoring in the lowest 25 percent on the pre-test, students in ESL programs and in

bilingual programs using the Spanish version of the Success for All curriculum, and minority students” (p. 116).

School Reform Support and Funding

After *A Nation at Risk*, the desire to improve student achievement and schools forced educators to analyze and evaluate schools in order to develop strategies to improve academic outcomes. Sizer (1984), indicated that “he was puzzled over the obvious inefficiencies of the basic structure of schools” (p. 222). Sizer (1984) also indicated that “piecemeal reform is no reform” (p. 227). “As dissatisfaction with American education in general, and the effectiveness of federal compensatory education programs in particular, rose after the early 1980s, both Democratic and Republican leaders sought to devise and promote comprehensive school reform efforts” (Gordon, 2003p. 129; Ravid, 2000).

We have learned that Seymour Sarason’s critique of 1960s school reform efforts was right, that everything of importance in a school affects everything else of importance; that narrow, piecemeal reform, of whatever power, almost inevitably is crushed by the power of existing routines (Keefe et al., 1992, p. 28). The Carnegie Foundation (1988) indicated that “we propose a comprehensive program, one that puts together, as a single strategy, the best practices we observed” (p. xv). As a result of these findings and beliefs regarding piecemeal reform, educators began to look more closely at school-wide reform approaches as a strategy to improve schools.

Comprehensive school reform (CSR) is one of the more popular terms used to describe the improvement of entire schools. Comprehensive school reform has been defined and described in many different ways. According to Protheroe (1998), “Comprehensive school improvement is based on reorganizing and revitalizing the entire

school, rather than focusing on specific student populations or programs” (p. v).

According to Teaching (1988), “Comprehensive school reform should consist of...effective common-sense procedures-from early intervention, to smaller schools, to flexible scheduling arrangements-should not be introduced in piecemeal fashion” (p. xvii). “In comprehensive schoolwide reform, all factors with the potential to affect teaching and learning are addressed in a coherent and comprehensive way” (Protheroe, 1998, p. 7).

Comprehensive school reform focuses on reorganizing and revitalizing the entire school rather than on isolated, piecemeal reforms. It provides teachers with training, materials, and support they need to help students reach challenging academic standards. In comprehensive school reform, a “design” becomes the basis for all operations and activities within the school, and provides a vision that helps the school focus its efforts, engage teachers in their work, and build strong parent and community support (N. A. Schools, 1998, p. 7).

An increase in the school-wide reform strategy occurred after 1994. “In 1994, the U.S. Congress made it easier for schools to organize themselves in this way. For the first time, schools with student poverty rates as low as 50 percent could use Title 1 funds to improve the entire school” (Herman, 1999, p. 1). Congress made it even easier to use school-wide reform as a strategy to improve schools in 1997. In 1997, “Congress authorized an additional \$145 million per year to help low-performing (mostly Title 1) schools raise student achievement by adopting “research-based, school-wide approaches” (Herman, 1999, p. 1). As a result of the new Title 1 legislation and additional funding,

“School-wide reform is an increasingly popular school improvement strategy, especially for low-performing, high-poverty schools” (Herman, 1999, p. 1).

Years of experience in reform shows that a fragmented approach to school improvement does not result in the significant increases in student achievement that schools aim for and the public wants. In addition, experiences shows that success depends on school communities deciding what students should know and be able to do, and then working to implement focused, researched-based improvement efforts (N. A. Schools, 1998, p. 7). “Based on research and experience, advocates of comprehensive school-wide reform believe that it holds the potential for raising student achievement not only for disadvantaged students, but for all students” (Protheroe, 1998, p. 7).

Baltimore City Public Schools Study

The Baltimore City Public Schools study is important because it was one of the first studies conducted to evaluate SFA. This study also investigates the effects of SFA on reading scores. The results of this study also provide evidence that indicates SFA has improved student achievement.

In 1991, the Success for All Foundation (SFAF) completed an evaluation of the effectiveness of SFA in the Baltimore City School District in Maryland. During this evaluation, five experimental (SFA) schools were matched with the five control (non-SFA) schools with similar demographics in Baltimore City. “Assessments of reading proficiency were individually administered to students by specially trained students from local colleges who were unaware of the study hypotheses or of the schools’ treatment status, and other data were obtained from school records” (Madden, 1993, p. 132). The

Woodcock Language Proficiency Battery and the Durrell Analysis of Reading Difficulty were used to assess students.

Within each matched school, students were individually matched on standardized achievement test scores from the Spring before implementation began. Only students in experimental and control schools who have been in their respective schools since first grade (or earlier) were included in this analysis (Madden, 1993, p. 131).

The schools were evaluated using a longitudinal study method over a period of three school years and analyzed using multivariate and univariate analyses. The students' reading levels were measured using Letter-Word, Word Attack and Oral Reading assessments.

The reading test data were analyzed using multivariate analyses of variance (MANOVAs), with the pretests (standard scores) as covariates and raw scores on the three reading scales as dependent measures. The MANOVAs produced Wilks's lambda statistics and tests of significance that indicate the program effect on the general "reading" factor. Following the multivariate analyses, univariate analyses of covariance (ANCOVAs) were computed for each dependent measure separately. Univariate analyses should be interpreted cautiously if multivariate analyses are not statistically significant at $p < .10$ or better (Madden, 1993, p. 132).

For first graders in the SFA schools, the multivariate analyses were statistically significant for all five schools. The univariate analyses for first graders were statistically significant for three out of five schools in Letter-Word, one out of five schools in Oral Reading and all five schools in Word Attack. For second graders in the SFA schools, the

multivariate analyses were statistically significant for four out of five schools. The univariate analyses for second graders were statistically significant for all five schools in Letter-Word, all five schools in Oral Reading and four out of five schools in Word Attack. For third graders in the SFA schools, the multivariate analyses were statistically significant for all five schools. The univariate analyses for third graders were statistically significant for four out of five schools in Letter-Word, all five schools in Oral Reading and four out of five schools in Word Attack.

The first-year assessment revealed substantially higher student performance on measures of language development in preschool and kindergarten and on measures of reading in Grades 1-3 compared to students in a matched school. Reading gains were especially large for students who had been in the lowest 25% of their grade on pretests: For these students, effect sizes¹ averaged + .80 on individually administered measures (Madden, 1993, p. 125).

Charleston County Schools Study

The Charleston County Schools study is important because it uses standardized test results to measure outcomes and student achievement. This study also investigated the effects of SFA on standardized tests. In addition, the results of this study provided mixed results regarding the effects of SFA on student achievement.

“The SFA program was introduced to the Charleston County School District (CCSD) in South Carolina in the 1988-1989 school year as part of a broad plan to improve the educational outcomes of the district schools” (Jones, Gottfredson, & Gottfredson, 1997, p. 648). The program developers conducted a study in which an experimental (Success for All) school and a control (non-Success for All) school were

used. “The two schools were chosen to be similar to each other in demographics and history of performance on district standardized tests” (Jones et al., 1997, p. 650). The grade levels were divided into cohorts and standardized tests and Success for All assessments were used to measure outcomes. The Charleston County School District wanted to use standardized tests to assess their schools’ improvement; therefore, they used the Metropolitan Readiness Test (MRT), the Stanford Achievement Test (SAT), the Cognitive Skills Assessment Battery and the Basic Skills Assessment Program (BSAP). SFA used the Woodcock Reading Mastery Tests-Revised, the Merrill Language Screening Test, the Test of Language Development (TOLD), and the Durrell Test of Reading Difficulty.

The schools were evaluated using a longitudinal study method over a period of three school years and analyzed using multivariate and univariate analyses. For each cohort, adjusted means, effect sizes, and standardized regression coefficients were calculated for each outcome measure. The adjusted means for the treatment and comparison groups were calculated using ANCOVA. The standardized regression coefficients (betas) were calculated from multiple regression analysis. These analyses paralleled the ANCOVAs, with SFA/comparison school status as the independent variable and fall first grade CSAB pretest score as a covariate (entered in the first step) for Cohort 1 and Cohorts 2 and 3. MANCOVAs were also run with the same covariates as a way to determine whether the univariate results should be regarded as simply results of Type I error (Jones et al., 1997, p. 656).

The following are results of the MANCOVA analyses for Cohorts 1, 2 and 3. For Cohort 1-Year 1, the Stanford Achievement Test Reading scores were statistically

significant and the Stanford Achievement Test Math scores were statistically significant. All other measures for Cohort 1 –Year 1, 2 and 3 were statistically insignificant or not applicable. For Cohort 2-Year 1, the Letter-Word Identification, Word Attack, Auditory, Visual Matching and Sentence Imitation scores were statistically significant. All other measures for Cohort 2 –Year 1, 2 and 3 were statistically insignificant or not applicable. For Cohort 3-Year 2, the Letter-Word Identification and Word Attack scores were statistically significant. All other measures for Cohort 3 –Year 2 and 3 were statistically insignificant or not applicable.

Table 2 presents the percentages of students in the Charleston County School District SFA schools that achieved the district standards on the Basic Skills Assessment Program for grades 1, 2 and 3 in 1989, 1990, 1991 and 1992 respectively.

Table 2.

Percentage of Charleston County School District SFA Schools Students Achieving the Basic Skills Assessment Program District Standards from 1989-1992

Grade Level	Year			
	1989	1990	1991	1992
1	95	71	72	70
2	87	73	60	61
3	85	82	90	67

(Jones et al., 1997, p. 650)

The results of this evaluation are mixed; some of the effects were significant and positive, others were nonsignificant, and still others were significant and negative. The main finding was that the kindergarten program produced large positive effects for the

SFA students. The positive effects were not found in the later grades, however, this finding is consistent with the independent evaluation of the Baltimore City SFA schools (Venezky), which found positive effects for the pre-K to first grade programs but little effect for the later grades (Grades 2 through 5) (Jones et al., 1997, p. 659). “The conclusions that can be drawn from the present study concur with those suggested by Smith, Ross, and Nunnery (1997) and Nunnery et al. (1997), who found that the level of implementation of the SFA program varies and that the effects of the program depend on the level of implementation” (Jones et al., 1997, p. 667).

New American Schools Study

The New American Schools study is important because the developers of New American Schools did not conduct it and it revealed some important finding regarding the whole-school reform models that were evaluated. Although this study does not focus specifically on SFA, 26.56% (17) of the schools in this study were SFA schools and 23.90% (913) of the student scores were from SFA schools. The results of this study indicate that New American Schools has not improved student achievement.

“New American Schools (NAS) is a nonprofit, nonpartisan organization that was founded in 1991, by business leaders who wanted to invest in improving the quality of public education in this country” (N. A. Schools, 1998, p. 3). Berends, Bodilly, & Kirby (2002) stated, “Spurred by the piecemeal approach to school reform that had produced little change in the nation’s test scores, New American Schools (NAS) launched its efforts for whole-school reform in 1991” (p. 1). The New American Schools joined a host of organizations attempting to develop whole-school reform models. According to Berends et al., (2002), “This initiative was based on the premise that high-quality schools

are established with external providers (design teams), providing assistance to schools for implementing designs” (p. 1). Design teams provided professional development and technical assistance to help school communities effectively implement the design’s curriculum and the management of the design. According to Berends et al., (2002), “The mission of NAS was to help schools and districts significantly raise the achievement of large numbers of students with whole-school designs and the assistance design teams provide during the implementation process” (p. 1).

Berends et al., (2002), indicated that, “NAS approached RAND to assess...document and analyze the conditions under which NAS made progress towards its goals of wide spread use and implementation of its designs and improved student performance associated with that use” (p. xvi). RAND conducted several studies of the New American Schools from 1992-2000 to evaluate the progress of the New American Schools. These studies resulted in a book titled, *Facing the Challenges of Whole-School Reform: New American Schools After a Decade*.

Over this time period, RAND’s program of studies included a longitudinal study that examined implementation and performance changes across the entire group of schools, implementing designs in the partner jurisdictions; case studies of schools; and a classroom study of implementation and performance in a specific district (Berends, Bodilly et al., 2002, p. 24). The New American Schools provided services for several whole-school reform models, including SFA. Data was collected from 184 schools in districts throughout the United States during the evaluation period.

The Challenges of Conflicting School Reforms: Effects of New American Schools in a High Poverty District study analyzed the implementation of the New

American Schools in San Antonio, Texas and its effects on teaching and student achievement. The New American Schools' designs were first implemented in San Antonio in the 1996-1997 school year. This study was conducted using data from the 1997-1998 school year. "During site visits and classroom observations in the Spring of 1998, design implementation was clearly under way in many schools" (Berends, Chun, Schuyler, Stockly, & Briggs, 2002, p. 115). The NAS designs that were analyzed during the study are Co-NECT Schools (CON), Expeditionary Learning Outward Bound (EL), Modern Read Schoolhouse (MRSH) and Roots & Wings (RW). The San Antonio study asked the following research questions:

(1) Do the NAS designs extend beyond changes in school organization and governance and permeate classrooms? (2) Do NAS teachers and students interact with each other and subject materials in ways that reflect the innovative curricular and instructional approaches of the design teams? (3) What factors at the district, school, and classroom level are related to implementation of designs and changes in classroom instruction (Berends, Bodilly et al., 2002, p. 25)?

"The press to increase student achievement and improve test scores in San Antonio schools was clearly evident during the time of our study" (Berends, Bodilly et al., 2002, p. 101). The Texas Assessment of Academic Skills (TAAS) was used to measure student achievement in this study.

Regarding the San Antonio study, Berends, Chun, Schuyler, Stockly, & Briggs (2002) stated, "Prior to estimating the full reading and mathematic models, a two-level ANOVA regression was used to estimate within- and between-classroom variance in the district for both TAAS reading and mathematics scores" (p. 126). The models included

independent variables from the following categories: student characteristics, teacher background, classroom characteristics and school characteristics.

The results of the ANOVA regarding reading scores revealed that 83.9% of the total variance in student achievement lies within classrooms and 16.9% of the total variance in student achievement lies between classrooms. The results of the district level models regarding reading scores revealed that 46.85% of the variance within the classrooms can be explained by the model and 66.96% of the variance between classrooms can be explained by the model. Therefore, 53.15% of the variance within the classrooms could not be explained by the model and 43.04% of the variance between classrooms could not be explained by the model. According to Berends, Chun et al. (2002), "These findings concur with similar analyses of student scores in the context of classrooms and schools (see Lee et al., 1998; Gamoran, 1992; Lee and Bryk, 1989)" (p. 127). Berends, Chun et al. (2002) also stated, "If most of the differences in student outcomes come from within classrooms and are student specific, rather than between classrooms, then school reform efforts which focus on instructional practices might not be able to produce desired results" (p. 127). "As expected because of the early stages of implementation, NAS designs had no significant effect on student achievement, as measured by TAAS reading and mathematics scores" (Berends, Chun et al., 2002, p. 133).

Singer Study

The Singer study is important because it evaluated the implementation of WSR in Abbott District schools and the impact of whole-school reform on academic achievement. Specifically, this study evaluates some of the same whole-school reform models that are

evaluated in this study. In addition, the Singer study used ESPA data to evaluate academic achievement.

The purpose of the Singer study is to investigate the academic achievement outcomes of whole-school reform implementation in the Abbott Districts in New Jersey. In the Singer study, the researcher used ESPA Language Arts Literacy, Mathematics and Science scores to evaluate academic achievement. The data collected and evaluated covered the 1999-2000, 2000-2001 and 2001-2002 school years of 292 schools. “The individual students in the population are not being studied directly. Instead, the mean fourth grade ESPA scores for general studies students in each school are studied” (Singer, 2002, p. 83).

Singer (2002) stated that ... “the use of ANOVA, MANCOVA and paired T-tests along with the appropriate correlations and distribution analyses are the more appropriate choices for statistical examination of the collected and selected data” (p. 83). “ANOVAs were performed for the ESPA LAL in all three years. The F ratios were found to be statistically significant in all three years...Following up with a Tukey HSD for each year, statistically significant mean differences were found” (Singer, 2002, p. 113). Statistically significant mean differences were found between SFA and Co-NECT schools and between SFA and School Development Program/Comer schools for the 1999 data. Statistically significant mean differences were found between SFA and School Development Program/Comer schools and between SFA and Coalition of Essential Schools for the 2000 data. Statistically significant mean differences were found between SFA and School Development Program/Comer schools, between SFA and Coalition of

Essential Schools and between SFA and Co-NECT schools for the 2001 data (see Appendix D).

Consistently across the three years studied, CoNECT, Comer, and Coalition of Essential Schools have shown the greatest gains and the higher scores. This held true for ESPAs in LAL, Math, and Science. SFA consistently had the lowest scores. Moreover, SFA remained, in 2001, the only WSR model whose mean scores for ESPA LAL were below the state proficient score of 200 (Singer, 2002, p. 137).

Current and Future Whole-School Reform

The history of the process of improving schools has taught us that this is a dynamic process. Therefore, we know that whole-school reform as we see it today will be different in the future. According to Marzano (2003), “we are in a new era of school reform” (p. 158). Marzano (2003) stated, “I believe that the new era of school reform is based on three principles that make it very different from its predecessors” (p. 158).

These three principles are:

Principle 1. The new era of school reform is based on the realization that reform is a highly contextualized phenomenon. Principle 2. The new era of school reform is characterized by a heavy emphasis on data. Principle 3. In the new era of school reform, change is approached on an incremental basis (Marzano, 2003, p. 158).

Basically, Marzano suggests that reform must be school specific, achievement must be measured by evaluating data and change must be implemented gradually.

Principles 1 and 2 have a significant impact on this study. Evidence of the impact of

Principle 1 is supported by evaluating the effectiveness of SFA as the appropriate whole-

school reform model for schools in the Urban Public Schools. Evidence of the impact of Principle 2 is supported by the use of data to evaluate academic achievement.

Conceptual Framework

As a result of this literature review, the logical framework analysis (LFA) was developed to provide the conceptual framework for this study. According to Anderson (1999), “One of the most powerful frameworks for investigating educational processes is the logical frame work analysis designed by the United States Agency for International Development” (p. 61). Table 3 is a chart that presents the conceptual framework for this study.

Table 3.

Conceptual Framework for This Study

Component	Description
Narrative Summary	Program goal: To produce students who are reading on or above grade level by the end of third grade.
Objectively Verifiable Indicators	Measure of goal achievement: SFA program student reading level assessments.
Means of Verification	Means of verification: SFA program implementation reports are satisfactory or better. Language Arts Literacy Assessments.
Important Assumptions	For achieving goal targets: Students are scoring poorly in Language Arts Literacy on the ESPA.

Table 3.

Conceptual Framework for This Study

Component	Description
Narrative Summary	Project Purpose: To produce 4th grade students who are proficient in Language Arts Literacy on the ESPA and NJ ASK4.
Objectively Verifiable Indicators	Conditions that will indicate that purpose has been achieved: No Child Left Behind Adequate Yearly Progress benchmarks achieved.
Means of Verification	Means of verification: ESPA and NJ ASK4 district and NJDOE reports.
Important Assumptions	For achieving purpose: The SFA program will be effective in producing students who are proficient in Language Arts Literacy.
Narrative Summary	Outputs: To provide a research based Language Arts Literacy program to increase student achievement.
Objectively Verifiable Indicators	Magnitude of Output: The program will provide SFA Language Arts Literacy instruction to approximately 7821 4 th Grade students from 2002-2004.

Table 3.

Conceptual Framework for This Study

Component	Description
Means of Verification	Means of verification: Enrollment statistics for SFA schools.
Important Assumptions	For achieving outputs: NJDOE continues to provide financial and technical support for the SFA program.
Narrative Summary	Inputs: Implementation to the SFA Whole-School Reform Model.
Objectively Verifiable Indicators	Implementation Target: SFA was implemented in 18 schools in the 1999-2000 school year and it was supported by Abbott funds.
Means of Verification	Means of verification: NJDOE approval and monitoring.
Important Assumptions	For providing inputs: NJDOE approved additional Reading Tutors and a Reading Facilitator.

Summary

This chapter reviewed relevant literature and provided a conceptual framework for this study. Literature regarding early education reform, the School Development Program/Comer model, the Coalition of Essential Schools model, the SFA model, the

Baltimore City Public Schools Study, the Charleston City Public Schools Study, the New American Schools Study and the Singer Study are reviewed in this chapter.

The existence of several whole-school reform models that use various strategies and methods to improve schools and student achievement were discussed. The importance of early education reform and the impact that it has had on whole-school reform was discussed. The School Development Program/Comer model was described and its relevance to this study was explained. The School Development Program/Comer model focuses on the personal, social and moral development of students. The efficacy of the School Development Program/Comer model and the impact that it has on student achievement was also discussed. The School Development Program/Comer model tended to produce positive effects on student achievement. The Coalition of Essential Schools model was described and its relevance to this study was explained. The Coalition of Essential Schools model focuses on making fundamental changes in the design, structure, leadership and community relationships of schools. The efficacy of the Coalition of Essential Schools model and the impact that it has on student achievement was also discussed. The Coalition of Essential Schools model tended to produce mixed effects on student achievement. The SFA model was described and its relevance to this study was explained. The SFA model focuses on highly structured reading instruction. The efficacy of the SFA model and the impact that it has on student achievement was also discussed. The SFA model produced positive and mixed effects on student achievement.

The Baltimore City Public Schools and the Charleston City Public Schools studies were early evaluations of the SFA model that were conducted by the developers. The Baltimore City Public Schools study indicated that SFA had a positive effect on student

achievement. The Charleston City Public Schools study indicated that SFA had mixed results that included both positive and negative effects on student achievement. The New American Schools study was not conducted by the developers and it indicated that SFA did not have positive effects on student achievement. The Singer study indicated that SFA had positive effects on student achievement. However, the Singer study also revealed that the SFA whole-school reform model schools were the only schools that had a mean Language Arts Literacy score that was below the passing score of 200 according to the 2001 data. A chart presenting an analysis of these studies can be found in Appendix H.

The conceptual framework for this study was also presented in this chapter. In addition, a table was also provided regarding the conceptual framework for this study.

CHAPTER III

METHODOLOGY

Introduction

This chapter explains the research design of this study and the methods employed to conduct the study. The first section explains the ESPA and the NJ ASK4, the second section explains the methods of data collection and the third section explains the methods of data analyses.

ESPA/NJ ASK4

“In May 1996, the New Jersey Board of Education adopted the Core Curriculum Content Standards (CCCS) which enumerated what all New Jersey students should know and be able to do by the end of the fourth and eighth grades, and upon completion of a New Jersey public school education” (N. J. D. O. Education, 2004b). “Deemed an early warning test, the Elementary School Proficiency Assessment (ESPA) (1997-2002) was administered to determine how well fourth-grade students were progressing toward acquiring skills needed to graduate from high school” (N. J. D. O. Education, 2004a). “In 2003, the NJ ASK program replaced the Elementary School Proficiency Assessment (ESPA) ... with a comprehensive, multi-grade assessment program for both third and fourth graders, as required by the No Child Left Behind Act of 2001” (ETS, 2004). “Along with other indicators of student progress, the results of the elementary level assessments are intended to be used to identify students who need additional instructional support in order to reach the CCCS” (N. J. D. O. Education, 2004a). “The testing times and test formats are similar to those for ESPA, including open-ended items for language

arts literacy (LAL) and calculator use and open-ended items for mathematics” (ETS, 2004).

Description of Sample/Population

The population studied in this research included the Urban Public School students who completed the ESPA during the 2001-2002 school year and NJ ASK4 during the 2002-2003 and the 2003-2004 school years. The population included 7,898 students from 48 schools. There were 18 SFA schools in the Urban Public Schools that administered the ESPA and NJ ASK4. Only general education scores were used during this study. The data used was analyzed based on School Group, WSR model, gender and ethnicity. It is important to note that the Urban Public Schools are an Abbott district because this helps to define the characteristics of the population that is studied.

Data Collection

The researcher collected ESPA and NJ ASK4 student, school and district data from the Urban Public Schools and the New Jersey Department of Education website.

The following data was obtained from the New Jersey Department of Education:

1. The names of the Urban Public Schools that have the SFA whole-school reform model.
2. The whole-school reform model cohort designation for SFA schools in the Urban Public Schools. Schools are designated cohorts 1, 2, 2-midyear, 3 and 3-midyear based on the year that they implemented their whole-school reform model.

A letter of request was sent to the Superintendent of the Urban Public Schools to obtain ESPA and NJ ASK4 data. A letter of approval was received from the Superintendent of

the Urban Public Schools granting permission to obtain the assessment data. The following data was obtained from the Urban Public Schools:

1. The names of the Urban Public Schools and the School Groups (School Group) to which they are assigned.
2. The ESPA individual student data for Grade 4 Urban Public Schools students for the 2001-2002 school year.
3. The NJ ASK4 individual student data for Grade 4 Urban Public Schools students for the 2002-2003 and the 2003-2004 school years.

Data Analysis

The purpose of this research design is to determine the effects of the SFA whole-school reform model on Language Arts Literacy and Reading Cluster mean scores. The research design was formulated to answer the following twelve research questions:

1. Are there significant differences among the cohorts when comparing Language Arts Literacy mean scores from 2002-2004?
2. Are there significant differences among the cohorts when comparing Reading Cluster mean scores from 2002-2004?
3. Are there significant differences between School Group A and other School Groups when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
4. Are there significant differences between School Group A and other School Groups when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?

5. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Language Arts Literacy mean scores from 2002-2003, 2003-2004 and 2002-2004?
6. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Reading Cluster mean scores from 2002-2003, 2003-2004 and 2002-2004?
7. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
8. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?
9. Are gender, ethnicity and the WSR model good predictors of Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
10. Are gender, ethnicity and the WSR model good predictors of Reading Cluster mean scores for school years 2002, 2003 and 2004?
11. Are gender and ethnicity good predictors of Non-SFA Schools Reading Cluster mean scores for school years 2002, 2003 and 2004?
12. Are gender and ethnicity good predictors of SFA Schools Reading Cluster mean scores for school years 2002, 2003 and 2004??

The secondary test data collected from the Urban Public Schools and the New Jersey Department of Education website was analyzed using the Statistical Package for Social Sciences (SPSS) to conduct a study that includes cross-sectional and longitudinal

methods. “A cross-sectional study involves observations of a sample, or cross section, of a population or phenomenon that are made at one point in time” (Babbie, 2002, p. 96).

“In contrast to cross-sectional studies, a longitudinal study is designed to permit observations of the same phenomenon over an extended period” (Babbie, 2002, p. 97). T-tests and multiple regressions were used to analyze the test data.

T-tests were performed to answer research questions 1, 2, 3, 4, 5, 6, 7 and 8. “A t-test is a procedure used for comparing sample means to see if there is sufficient evidence to infer that the means of corresponding population distributions also differ” (George & Mallery, 2001, p. 122). Independent sample t-tests were performed to analyze the ESPA and the NJ ASK4 assessment data. According to George & Mallery (2001), the independent-samples t-test, compares the means of two different samples. The two samples share some variable of interest in common, but there is no overlap between membership of the two groups (p. 122). Independent sample t-tests were performed to compare the ESPA and the NJ ASK4 Language Arts Literacy mean scores and the Reading Cluster mean scores. Independent sample t-tests were performed to compare the ESPA and the NJ ASK4 Language Arts Literacy and the Reading Cluster mean scores of SFA and Non-SFA schools and School Group A and other School Groups for 2002, 2003 and 2004. Independent sample t-tests were also performed to compare Language Arts Literacy and the Reading Cluster mean scores for the district, SFA and Non-SFA schools from 2002 to 2004.

Multiple regressions were performed to answer research questions 9, 10, 11 and 12. “When two or more variables are used to predict one criterion variable, the multiple regression procedure is employed” (Ravid, 2000, p. 181). Multiple regressions were

performed to determine if there was a gender, ethnicity and WSR model effect on District schools' ESPA and the NJ ASK4 Language Arts Literacy mean scores for 2002, 2003 and 2004. Multiple regressions were performed to determine if there a was gender, ethnicity and WSR model effect on District schools' ESPA and the NJ ASK4 Reading Cluster mean scores for 2002, 2003 and 2004. Multiple regressions were performed to determine if there was a gender and ethnicity effect on Non-SFA schools' ESPA and the NJ ASK4 Reading Cluster mean scores for 2002, 2003 and 2004. Multiple regressions were performed to determine if there was a gender and ethnicity effect on SFA schools' ESPA and the NJ ASK4 Reading Cluster mean scores for 2002, 2003 and 2004.

Summary

This chapter explained the ESPA and the NJ ASK4 assessments. These assessments are administered to fourth grade students in New Jersey to measure student achievement and to serve as an early warning for students who are not acquiring the skills required by the New Jersey Core Curriculum Content Standards. The chapter also contained the description of the sample, methods of data collection, methods of data analysis and the research questions. This chapter indicated that ESPA and the NJ ASK4 assessment data was collected regarding the fourth grade students in the Urban Public Schools. ESPA and the NJ ASK4 assessment data was also collected from the New Jersey Department of Education website. This chapter also described the research design of the study and the methods employed to complete the study. The study used cross-sectional and longitudinal methods and analyzed secondary data by performing t-tests and multiple regression analyses.

CHAPTER IV

DATA ANALYSIS

Introduction

This chapter provides a summary of results obtained from the various analyses conducted for this study using SPSS (Statistical Package for the Social Sciences). In the beginning of this chapter, the purpose of this study and the research questions are restated. The remaining sections of the chapter are divided based on the analyses performed to answer the research questions.

The purpose of this study is to investigate the impact of the SFA whole-school reform model on student achievement as measured by the ESPA and NJ ASK4 in the Urban Public Schools. The following research questions were explored during this study:

1. Are there significant differences among the cohorts when comparing Language Arts Literacy mean scores from 2002-2004?
2. Are there significant differences among the cohorts when comparing Reading Cluster mean scores from 2002-2004?
3. Are there significant differences between School Group A and other School Groups when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
4. Are there significant differences between School Group A and other School Groups when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?

5. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Language Arts Literacy mean scores from 2002-2003, 2003-2004 and 2002-2004?
6. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Reading Cluster mean scores from 2002-2003, 2003-2004 and 2002-2004?
7. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
8. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?
9. Are gender, ethnicity and the WSR model good predictors of Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
10. Are gender, ethnicity and the WSR model good predictors of Reading Cluster mean scores for school years 2002, 2003 and 2004?
11. Are gender and ethnicity good predictors of Non-SFA Schools Reading Cluster mean scores for school years 2002, 2003 and 2004?
12. Are gender and ethnicity good predictors of SFA Schools Reading Cluster mean scores for school years 2002, 2003 and 2004?

The remainder of this chapter is divided into sections based on the research questions and the results of the appropriate statistical analyses. T-tests and multiple regressions were conducted to analyze ESPA and the NJ ASK4 assessment data.

Research Question 1

Are there significant differences among the cohorts when comparing Language Arts Literacy mean scores from 2002-2004?

Independent sample t-tests were conducted to determine the differences between District Language Arts Literacy means for 2002-2003, 2003-2004 and 2002-2004. The results from the t-tests can be found in Tables 4, 5 and 6.

The 2002 mean was 207.60 and the 2003 mean was 205.59. The difference between the 2002 and 2003 mean was 2.016 and it was significant with a $t = 3.377$ and $p < .001$. The 2004 mean was 208.60. The difference between the 2003 and 2004 mean was -3.015 and it was significant with a $t = -4.479$ and $p < .000$. The difference between the 2002 and the 2004 mean was -.999 and it was not significant with a $t = -1.572$ and $p < .116$.

Table 4.

T-Test for District Language Arts Literacy Means for 2002-2003

Variable	N	Year	M	SD	t	df	Sig.
LANGSCSC	2910	2002	207.60	21.494			
LANGSCSC	2598	2003	205.59	22.794			
Mean Difference			2.016		3.377	5506	.001**

Table 5.

T-Test for District Language Arts Literacy Means for 2003-2004

Variable	N	Year	M	SD	t	df	Sig.
----------	---	------	---	----	---	----	------

LANGSCSC	2598	2003	205.59	22.794			
LANGSCSC	2311	2004	208.60	24.350			
Mean Difference			-3.015		-4.479	4907	.000**

Table 6.

T-Test for District Language Arts Literacy Means for 2002-2004

Variable	N	Year	M	SD	t	df	Sig.
LANGSCSC	2910	2002	207.60	21.494			
LANGSCSC	2311	2004	208.60	24.350			
Mean Difference			-.999		-1.572	5219	.116

This research question was concerned with significant differences on a district level when comparing Language Arts Literacy mean scores from 2002 to 2004. The analyses attempted to determine if WSR had an impact on student achievement. The analyses for this question resulted in a significant decrease in the district Language Arts Literacy mean score from 2002 to 2003 and a significant increase in the District Language Arts Literacy mean score from 2003 to 2004. However, the increase in Language Arts Literacy mean score from 2002 to 2004 was not significant.

The Urban Public Schools began implementing WSR in 1999. Since 1999, Language Arts Literacy mean scores have increased each year except for 2003 when the mean score decreased significantly. However, The Language Arts Literacy mean scores recovered in 2004, with a significant increase. Therefore, during this study from 2002 to

2004, the positive trend has continued; however, the trend was not significant. The 2003-2004 increase is consistent with year to year increases that have occurred during 5 out of the 6 years of WSR implementation. These trends and the increases suggest that WSR is continuing to have a positive impact on student achievement as measured by ESPA and NJASK Language Arts Literacy mean scores.

Research Question 2

Are there significant differences among the cohorts when comparing Reading Cluster mean scores from 2002-2004?

Independent sample t-tests were conducted to determine the differences between District Reading Cluster mean scores for 2002-2003, 2003-2004 and 2002-2004. The results from the t-tests can be found in Tables 7, 8 and 9.

The 2002 mean was 9.998 and the 2003 mean was 9.243. The difference between the 2002 and 2003 mean was .7554 and it was significant with a $t = 7.035$ and $p < .000$. The 2004 mean was 11.388. The difference between the 2003 and 2004 mean was -2.1454 and it was significant with a $t = -17.618$ and $p < .000$. The difference between the 2002 and 2004 mean was -1.3900 and it was significant with a $t = -12.863$ and $p < .000$.

Table 7.

T-Test for District Reading Cluster Means for 2002-2003

Variable	N	Year	M	SD	t	df	Sig.
READING	2910	2002	9.998	3.6292			
READING	2608	2003	9.243	4.3418			
Mean Difference			.7554		7.035	5516	.000**

Table 8.

T-Test for District Reading Cluster Means for 2003-2004

Variable	N	Year	M	SD	t	df	Sig.
READING	2608	2003	9.243	4.3418			
READING	2311	2004	11.388	4.1712			
Mean Difference			-2.1454		-17.618	4917	.000**

Table 9.

T-Test for District Reading Cluster Means for 2002-2004

Variable	N	Year	M	SD	t	df	Sig.
READING	2910	2002	9.998	3.6292			
READING	2311	2004	11.388	4.1712			
Mean Difference			-1.3900		-12.863	5219	.000**

This research question was concerned with significant differences on a district level when comparing Reading Cluster mean scores from 2002 to 2004. The analyses attempted to determine if WSR had an impact on Reading achievement. The analyses for this question resulted in a significant decrease in the District Reading Cluster mean score from 2002 to 2003 and a significant increase in the District Reading Cluster mean score from 2003 to 2004. In addition, the increase in Reading Cluster mean scores from 2002 to 2004 was significant.

The Reading Cluster mean score decrease is consistent with the Language Arts Literacy 2002-2003 decrease and the increase is consistent with the Language Arts Literacy increase for 2003 to 2004. The Reading Cluster mean score increase for 2002 to 2004 is consistent with the Language Arts Literacy mean score increase for 2002 to 2004; however, the Reading Cluster mean score exhibited a positive trend that was significant and the Language Arts Literacy mean score was not significant. The analysis also indicated that WSR is having a greater impact on Reading Cluster than it is having on the comprehensive Language Arts Literacy mean scores. In addition, the trend suggests that WSR is continuing to have a positive impact on student achievement as measured by ESPA and NJASK Reading Cluster mean scores.

Research Question 3

Are there significant differences between School Group A and other School Groups when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?

Independent sample t-tests were conducted to determine the Language Arts Literacy mean differences between School Group A and a combined group of schools from School Groups B, C, and D for 2002, 2003 and 2004. The Urban School District is divided into 5 School Groups and, in School Group A, 10 out of 12 schools are SFA schools. In School Group A, all of the schools where the ESPA or the NJ ASK4 is conducted implement the SFA WSR model. In the combination of Groups B, C and D only 8 schools implement SFA. The results from the t-tests can be found in Tables 10, 11 and 12.

The 2002 School Group A mean was 200.35 and the mean for the 2002 combination of Groups B, C, and D mean was 210.24. The mean difference between the 2002 School Group A and the 2002 combination of School Groups B, C, and D and was 9.887 and it was significant with a $t = 11.205$ and $p < .000$. The 2003 School Group A mean was 195.88 and the combination of 2003 School Groups B, C, and D mean was 209.84. The mean difference between the 2003 School Group A and the 2003 combination of School Groups B, C, and D was 13.958 and it was significant with a $t = 14.939$ and $p < .000$. The 2004 School Group A mean was 198.57 and the combination of 2004 School Groups B, C, and D mean was 212.68. The mean difference between the 2004 School Group A and the 2004 combination of School Groups B, C, and D and was 14.115 and it was significant with a $t = 13.090$ and $p < .000$.

Table 10.

T-tests for School Groups Language Arts Literacy for 2002

Variable	N	Year	Group	M	SD	t	df	Sig.
LANGSCSC	2134	2002	B, C & D	210.24	21.287			
LANGSCSC	776	2002	A	200.35	20.378			
Mean Difference				9.887		11.205	2908	.000**

Table 11

T-tests for School Groups Language Arts Literacy for 2003

Variable	N	Year	Group	M	SD	t	df	Sig.
LANGSCSC	1814	2003	B, C & D	209.84	22.105			

LANGSCSC	786	2003	A	195.88	21.354				
Mean Difference				13.958		14.939	2598	.000**	

Table 12.

T-tests for School Groups Language Arts Literacy for 2004

Variable	N	Year	Group	M	SD	t	df	Sig.
LANGSCSC	1643	2004	B, C & D	212.68	23.419			
LANGSCSC	668	2004	A	198.57	23.694			
Mean Difference				14.115		13.090	2309	.000**

This research question was concerned with significant Language Arts Literacy differences between School Group A and the combination of the other School Groups. The analysis attempted to determine the impact of the majority presence of the SFA schools in School Group A and their performance as compared to all other schools with other WSR models, including 8 SFA schools also. The analysis for this question resulted in a significant difference in Language Arts Literacy mean scores between School Group A and other School Groups favoring the other School Groups in 2002, 2003 and 2004.

There were significant mean differences between Group A and the other Groups during each of the three years of the study. This indicates that Group A schools are achieving at a lower level than schools from the other groups.

Research Question 4

Are there significant differences between School Group A and other School Groups when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?

Independent sample t-tests were conducted to determine the Reading Cluster mean differences between School Group A schools and a combined group of schools from School Groups B, C, and D for 2002, 2003 and 2004. The results from the t-test can be found in Tables 13, 14 and 15.

The 2002 Group A mean was 9.097 and the 2002 mean for the combination of School Groups B, C and D was 10.326. The mean difference between the 2002 Group A and the 2002 combination of School Groups B, C and D was 1.2284 and it was significant with a $t = 8.165$ and $p < .000$. The 2003 Group A mean of 7.530 and the combination of 2003 School Groups B, C and D mean was 9.988. The mean difference between the 2003 Group A and the 2003 combination of School Groups B, C and D was 2.4578 and it was significant with a $t = 13.761$ and $p < .000$. The 2004 Group A mean of 9.80 and the combination of 2004 School Groups B, C and D mean was 12.03. The mean difference between the 2004 Group A and the 2004 combination of School Groups B, C and D was 2.230 and it was significant with a $t = 12.009$ and $p < .000$.

Table 13.

T-tests for School Groups Reading Cluster for 2002

Variable	N	Year	Group	M	SD	t	df	Sig.
READING	2134	2002	B, C & D	10.326	3.3797			
READING	776	2002	A	9.097	3.6620			

Mean Difference	1.2284	8.165	2908	.000**
-----------------	--------	-------	------	--------

Table 14.

T-tests for School Groups Reading Cluster for 2003

Variable	N	Year	Group	M	SD	t	df	Sig.
READING	1817	2003	B, C & D	9.988	4.3450			
READING	791	2003	A	7.530	3.8205			
Mean Difference				2.4578		13.761	2606	.000**

Table 15.

T-tests for School Groups Reading Cluster for 2004

Variable	N	Year	Group	M	SD	t	df	Sig.
READING	1643	2004	B, C & D	12.03	4.086			
READING	668	2004	A	9.80	3.953			
Mean Difference				2.230		12.009	2309	.000**

This research question was concerned with significant Reading Cluster differences between School Group A and the combination of the other School Groups. The analysis attempted to determine impact on student achievement of the majority presence of the SFA schools in School Group A and the other WSR models, including SFA schools, in all of the other School Groups combined. The analysis for this question resulted in a significant difference in Reading Cluster mean scores between School Group A and other School Groups, favoring the other School Groups in 2002, 2003 and 2004.

There were significant mean differences between Groups A and the other Groups during each of the three years of the study. This suggests that Group A schools are achieving at a lower level than schools from the other groups. These findings are consistent with the finding from the Language Arts Literacy results. However, considering the fact that SFA is principally a reading program, these results are surprising because of the expectation that School Group A schools might compare more favorably to the other Groups when analyzing by Reading Cluster.

Research Question 5

Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Language Arts Literacy mean scores from 2002-2003, 2003-2004 and 2002-2004?

Independent sample t-tests were conducted to determine the differences between the group of Non-SFA schools Language Arts Literacy means for 2002-2003, 2003-2004 and 2002-2004. The results from the t-tests can be found in Tables 16, 17 and 18.

The 2002 mean was 210.77 and the 2003 mean was 210.49. The difference between the 2002 and 2003 mean was .282 and it was not significant with a $t = .363$ and $p < .716$. The 2004 mean was 212.80. The difference between the 2003 and 2004 mean was -2.310 and it was significant with a $t = -2.708$ and $p < .007$. The difference between the 2002 and the 2004 mean was -2.028 and it was significant with a $t = -2.553$ and $p < .011$.

Table 16.

T-Test for Non-SFA Schools Language Arts Literacy Means for 2002-2003

Variable	N	Year	M	SD	t	df	Sig.
----------	---	------	---	----	---	----	------

LANGSCSC	1811	2002	210.77	21.678			
LANGSCSC	1440	2003	210.49	22.303			
Mean Difference			.282		.363	3249	.716

Table 17.

T-Test for Non-SFA Schools Language Arts Literacy Means for 2003-2004

Variable	N	Year	M	SD	t	df	Sig.
LANGSCSC	1440	2003	210.49	22.303			
LANGSCSC	1416	2004	212.80	23.284			
Mean Difference			-2.310		-2.708	2854	.007**

Table 18.

T-Test for Non-SFA Schools Language Arts Literacy Means for 2002-2004

Variable	N	Year	M	SD	t	df	Sig.
LANGSCSC	1811	2002	210.77	21.678			
LANGSCSC	1416	2004	212.80	23.284			
Mean Difference			-2.028		-2.553	3225	.011**

Independent sample t-tests were conducted to determine the differences between the group of SFA schools Language Arts Literacy means for 2002-2003, 2003-2004 and 2002-2004. The results from the t-tests can be found in Tables 19, 20 and 21.

The 2002 mean was 202.38 and the 2003 mean was 199.49. The difference between the 2002 and 2003 mean was 2.891 and it was significant with a $t = 3.258$ and $p < .001$. The 2004 mean was 201.96. The difference between the 2003 and 2004 mean was -2.469 and it was significant with a $t = -2.403$ and $p < .016$. The difference between the 2002 and the 2004 mean was .421 and it was not significant with a $t = .421$ and $p < .674$.

Table 19.

T-Test for SFA Schools Language Arts Literacy Means for 2002-2003

Variable	N	Year	M	SD	t	df	Sig.
LANGSCSC	1099	2002	202.38	20.136			
LANGSCSC	1158	2003	199.49	21.918			
Mean Difference			2.891		3.258	2255	.001**

Table 20.

T-Test for SFA Schools Language Arts Literacy Means for 2003-2004

Variable	N	Year	M	SD	t	df	Sig.
LANGSCSC	1158	2003	199.49	21.918			
LANGSCSC	895	2004	201.96	24.530			
Mean Difference			-2.469		-2.403	2051	.016**

Table 21.

T-Test for SFA Schools Language Arts Literacy Means for 2002-2004

Variable	N	Year	M	SD	t	df	Sig.
----------	---	------	---	----	---	----	------

LANGSCSC	1099	2002	202.38	20.136		
LANGSCSC	895	2004	201.96	24.530		
Mean Difference			.421		.421	1992 .674

This research question was concerned with significant differences on a WSR level, when comparing Language Arts Literacy mean scores from 2002 to 2004. The analyses attempted to determine if the WSR models had an impact on student achievement. The analyses for this question resulted in significant Language Arts Literacy increases for Non-SFA schools from 2003-2004 and 2002-2004. Also, the analyses for this question resulted in a significant Language Arts Literacy increase for SFA schools from 2003-2004. The analyses for this question resulted in a significant Language Arts Literacy decrease for SFA schools from 2002-2003.

The 2003-2004 Language Arts Literacy increase for Non-SFA schools is consistent, with year to year increases that have occurred during 5 out of the 6 years of the WSR implementation. This increase is also consistent with the 2003-2004 results from analysis for the district for Language Arts Literacy and Reading. The 2002-2004 increase is consistent with the 2002-2004 positive trend that resulted from the analysis for the district for Reading for 2002-2004. Therefore, during this study the 2003-2004 increase and the 2002 to 2004 positive trend has continued and it is significant. These trends and the increases suggest that WSR is continuing to have a positive impact on student achievement, as measured by ESPA and NJASK Language Arts Literacy mean scores.

The 2002-2003 Language Arts Literacy decrease for SFA schools is consistent with the results from the analysis for the district that resulted in a decrease for Language Arts Literacy for 2002-2003. The 2003-2004 increase is also consistent with the results from analyses for the district for Language Arts Literacy and Reading and Non-SFA schools for Language Arts Literacy for 2003-2004. During this study, the SFA schools Language Arts Literacy scores significantly increased during the last year of the study. The 2003-2004 increase is consistent with the 2003-2004 positive trends that resulted from the analysis for the district for Reading, Non-SFA schools for Language Arts Literacy and SFA schools for Language Arts Literacy for 2003-2004. Therefore, during this study, from 2003 to 2004, the increase has continued and it is significant. This increase suggests that SFA has had a positive impact on student achievement as measured by the 2003-2004 NJ ASK4 Language Arts Literacy mean scores.

Research Question 6

Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Reading Cluster mean scores from 2002-2003, 2003-2004 and 2002-2004?

Independent sample t-tests were conducted to determine the differences between the group of Non-SFA schools Reading Cluster mean scores for 2002-2003, 2003-2004 and 2002-2004. The results from the t-tests can be found in Tables 22, 23 and 24.

The 2002 mean was 10.42 and the 2003 mean was 10.11. The difference between the 2002 and 2003 mean was .305 and it was significant with a $t = 2.144$ and $p < .032$. The 2004 mean was 12.016. The difference between the 2003 and 2004 mean was -1.903

and it was significant with a $t = -12.008$ and $p < .000$. The difference between the 2002 and the 2004 mean was -1.598 and it was significant with a $t = -11.654$ and $p < .000$.

Table 22.

T-Test for Non-SFA Schools Reading Cluster Means for 2002-2003

Variable	N	Year	M	SD	t	df	Sig.
READING	1811	2002	10.42	3.707			
READING	1443	2003	10.11	4.404			
Mean Difference			.305		2.144	3252	.032*

Table 23.

T-Test for Non-SFA Schools Reading Cluster Means for 2003-2004

Variable	N	Year	M	SD	t	df	Sig.
READING	1443	2003	10.11	4.404			
READING	1416	2004	12.02	4.060			
Mean Difference			-1.903		-12.088	2857	.000**

Table 24.

T-Test for Non-SFA Schools Reading Cluster Means for 2002-2004

Variable	N	Year	M	SD	t	df	Sig.
READING	1811	2002	10.42	3.707			
READING	1416	2004	12.02	4.060			

Mean Difference	-1.598	-11.654 3225	.000**
-----------------	--------	--------------	--------

Independent sample t-tests were conducted to determine the differences between the group of SFA schools Language Arts Literacy means for 2002-2003, 2003-2004 and 2002-2004. The results from the t-tests can be found in Tables 25, 26 and 27.

The 2002 mean was 9.31 and the 2003 mean was 8.17. The difference between the 2002 and 2003 mean was 1.142 and it was significant with a $t = 7.294$ and $p < .000$. The 2004 mean was 10.39. The difference between the 2003 and 2004 mean was -2.230 and it was significant with a $t = -12.313$ and $p < .000$. The difference between the 2002 and the 2004 mean was -1.088 and it was significant with a $t = -6.445$ and $p < .000$.

Table 25.

T-Test for SFA Schools Reading Cluster Means for 2002-2003

Variable	N	Year	M	SD	t	df	Sig.
READING	1099	2002	9.31	3.387			
READING	1165	2003	8.17	4.012			
Mean Difference			1.142		7.294	2262	.000**

Table 26.

T-Test for SFA Schools Reading Cluster Means for 2003-2004

Variable	N	Year	M	SD	t	df	Sig.
READING	1165	2003	8.17	4.012			
READING	895	2004	10.39	4.154			

Mean Difference	-2.230	-12.313 2058	.000**
-----------------	--------	--------------	--------

Table 27.

T-Test for SFA Schools Reading Cluster Means for 2002-2004

Variable	N	Year	M	SD	t	df	Sig.
READING	1099	2002	9.31	3.387			
READING	895	2004	10.39	4.154			
Mean Difference			-1.088		-6.445	1992	.000**

This research question was concerned with significant differences on a WSR level when comparing Reading Cluster mean scores from 2002 to 2004. The analyses attempted to determine if WSR is having an impact on Reading achievement. The analyses for this question resulted in significant Reading Cluster increases for Non-SFA schools from 2003-2004 and 2002-2004 and a significant decrease for Non-SFA schools from 2002-2003. The analyses for this question also resulted in significant Reading Cluster increases for SFA schools from 2003-2004 and 2002-2004 and a significant Reading Cluster decrease for SFA schools from 2002-2003.

The 2002-2003 Reading Cluster decrease for Non-SFA schools is consistent with the results from the analyses for the district for Language Arts Literacy and Reading and for Non-SFA schools for Reading for 2002-2003. The 2003-2004 Reading Cluster increase for Non-SFA schools is consistent with year to year increases that have occurred during 5 out of the 6 years of WSR implementation. This increase is also consistent with the 2003-2004 results from analyses for the district for Language Arts Literacy and

Reading and Non-SFA schools for Language Arts Literacy. The 2002-2004 increase is consistent with the 2002-2004 positive trends that resulted from the analyses for the district for Reading and Non-SFA schools for Language Arts Literacy for 2002-2004. Therefore, during this study, the 2003-2004 increase and the 2002 to 2004 positive trend has continued and it is significant. The trend and the increase suggest that other WSR models are continuing to have a positive impact on student achievement, as measured by ESPA and NJASK4 Language Arts Literacy mean scores.

The 2002-2003 Reading Cluster decrease for SFA schools is consistent with the results from the analyses for the district for Language Arts Literacy and Reading for Language Arts Literacy for 2002-2003. The 2003-2004 increase is also consistent with the results from analyses for the district for Language Arts Literacy and Reading and Non-SFA schools for Language Arts Literacy for 2003-2004. The 2002-2004 increase is consistent with the 2002-2004 positive trend that resulted from analyses for the district for Reading and Non-SFA schools for Language Arts Literacy for 2002-2004. Therefore, during this study, the 2003-2004 increase and the 2002 to 2004 positive trend has continued and it is significant. These trends and the increases suggest that SFA is continuing to have a positive impact on student achievement, as measured by NJASK Reading Cluster mean scores.

A comparison between the research questions 5 and 6 revealed that Language Arts Literacy increases and decreases are the same for 2002-2003 and 2003-2004. However, for 2002-2004, SFA schools increased and Non-SFA schools did not. The Reading comparison revealed that Non-SFA increases or decreases were the same for

2002-2003, 2003-2004 and 2002-2004. The results of the comparison can be found in Table 28.

Table 28.

Non-SFA and SFA Increases and Decreases for Language Arts Literacy and Reading Cluster Means

Year		Non-SFA Schools	SFA Schools
2002-2003	LAL	Decrease	Decrease
2003-2004	LAL	Increase	Increase
2002-2004	LAL	No Change	Increase
2002-2003	Reading	Decrease	Decrease
2003-2004	Reading	Increase	Increase
2002-2004	Reading	Increase	Increase

Research Question 7

Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?

Independent sample t-tests were conducted to determine the Language Arts Literacy mean differences between the group of Non-SFA schools and the group of SFA schools for 2002, 2003 and 2004. The results from the t-test can be found in Tables 29, 30 and 31.

The group of Non-SFA schools mean score was 210.77 and the group of SFA schools mean score was 202.38 for 2002. The mean difference between the group of

Non-SFA schools and the group of SFA schools was -8.397 and it was significant with a $t = -10.403$ and $p < .000$. The group of Non-SFA schools mean score was 209.46 and the group of SFA schools mean score was 199.24 for 2003. The mean difference between the group of Non-SFA schools mean score and the group SFA schools was -10.221 and it was significant with a $t = -11.340$ and $p < .000$. The group of Non-SFA schools mean score was 212.68 and the group of SFA schools mean score was 201.96 for 2004. The mean difference between the group of Non-SFA schools mean score and the group of SFA schools mean scores was -10.714 and it was significant with a $t = -10.509$ and $p < .000$.

Table 29.

T-tests for Non-SFA and SFA Schools for Language Arts Literacy for 2002

Variable	N	Year	Group	M	SD	t	df	Sig.
LANGSCSC	1811	2002	Non-SFA	210.77	21.678			
LANGSCSC	1099	2002	SFA	202.38	20.136			
Mean Difference				-8.397		-10.403	2908	.000**

Table 30.

T-tests for Non-SFA and SFA Schools for Language Arts Literacy for 2003

Variable	N	Year	Group	M	SD	t	df	Sig.
LANGSCSC	1623	2003	Non-SFA	209.46	22.646			
LANGSCSC	977	2003	SFA	199.24	21.596			
Mean Difference				-10.221		-11.340	2598	.000**

Table 31.

T-tests for Non-SFA and SFA Schools for Language Arts Literacy for 2004

Variable	N	Year	Group	M	SD	t	df	Sig.
LANGSCSC	1432	2004	Non-SFA	212.68	23.236			
LANGSCSC	879	2004	SFA	201.96	24.673			
Mean Difference				-10.714		-10.509	2309	.000**

This research question was concerned with significant Language Arts Literacy differences between the group of Non-SFA schools and the group of SFA schools. The analyses attempted to determine the Language Arts Literacy differences between the group of Non-SFA schools and the group of SFA schools. The analyses for this question resulted in significant differences in Language Arts Literacy mean scores between the group of Non-SFA schools and the group of SFA schools for 2002, 2003 and 2004.

There were significant mean differences between the group of Non-SFA schools and the group of SFA schools during each of the three years of the study. This indicates that the group of SFA schools are achieving at a lower level than the group of Non-SFA schools. These findings are consistent with the findings from the Language Arts Literacy results for School Groups and WSR and Reading for School Groups.

Research Question 8

Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?

Independent sample t-tests were conducted to determine the Reading Cluster mean differences between the group of Non-SFA schools and the group of SFA schools for 2002, 2003 and 2004. The results from the t-test can be found in Tables 32, 33 and 34.

The group of Non-SFA schools mean score was 10.418 and the group of SFA schools mean score was 9.306 for 2002. The mean difference between the group of Non-SFA schools and the group of SFA schools was -1.1118 and it was significant with a $t = -8.100$ and $p < .000$. The group of Non-SFA schools mean score was 10.43 and the group of SFA schools mean score was 8.165 for 2003. The mean difference between the group of Non-SFA schools mean score and the group SFA schools was 1.9482 and it was significant with a $t = 11.684$ and $p < .000$. The group of Non-SFA schools mean scores was 12.02 and the group of SFA schools mean score was 10.39 for 2004. The mean difference between the group of Non-SFA schools and the group of SFA schools was 1.622 and it was not significant with a $t = 9.271$ and $p < .128$.

Table 32.

T-tests for Non-SFA and SFA Schools for Reading Cluster for 2002

Variable	N	Year	Group	M	SD	t	df	Sig.
READING	1811	2002	Non-SFA	10.418	3.7073			
READING	1099	2002	SFA	9.306	3.3866			
Mean Difference				-1.1118		-8.100	2908	.000**

Table 33.

T-tests for Non-SFA and SFA Schools for Reading Cluster for 2003

Variable	N	Year	Group	M	SD	t	df	Sig.
----------	---	------	-------	---	----	---	----	------

READING	1443	2003	Non-SFA	10.113	4.4041			
READING	1165	2003	SFA	8.165	4.0114			
Mean Difference				1.9482		11.684	2606	.000**

Table 34.

T-tests for Non-SFA and SFA Schools for Reading Cluster for 2004

Variable	N	Year	Group	M	SD	t	df	Sig.
READING	1416	2004	Non-SFA	12.02	4.060			
READING	895	2004	SFA	10.39	4.154			
Mean Difference				1.622		9.271	2309	.128

This research question was concerned with significant Reading Cluster differences between the group of Non-SFA schools and the group of SFA schools. The analyses attempted to determine the Reading Cluster mean differences between the group of Non-SFA schools and the group of SFA schools. The analyses for this question resulted in significant differences in Reading Cluster mean scores between the group of Non-SFA schools and the group of SFA schools favoring the Non-SFA schools in 2002 and 2003.

There were significant mean differences between the group of Non-SFA schools and the group of SFA schools during the first two years of the study. This indicates that the group SFA schools are achieving at a lower level than Non-SFA schools. However, considering the fact that SFA is principally a reading program, these results are surprising because of the expectation that the group of SFA schools might compare more favorably to the Non-SFA schools when analyzing by Reading Cluster.

Research Question 9

Are gender, ethnicity and the WSR model good predictors of Language Arts Literacy mean scores for school years 2002, 2003 and 2004?

Multiple regressions were conducted to determine the effects of gender (GENDER02), ethnicity (ETHNICITY02) and the WSR Model (WSRMODEL) on Language Arts Literacy scores (LANGSCSC02). The dependent variable was LANGSCSC02 and the independent variables were GENDER02, ETHNICITY02 and WSRMODEL. The results of the analysis can be found in Table 35.

The Regression analysis revealed a LANGSCSC02 mean of 207.62. The correlations revealed a significant Pearson $r = -.189$ between LANGSCSC02 and WSRMODEL, a significant Pearson $r = .185$ between LANGSCSC02 and GENDER02, and a significant Pearson $r = .147$ between LANGSCSC02 and ETHNICITY02. The R Square was computed for each model. The R Square identifies the proportion of the variance in the LANGSCSC02 mean score that can be attributed to the model. In Model 1 R Square = .036, Model 2 R Square = .071 and Model 3 R Square = .093. The ANOVA table revealed that Model 1 was significant at $F = 107.712$ ($df_1 = 2906$) and $p < .000$. The ANOVA table revealed that Model 2 was significant at $F = 110.361$ ($df_2 = 2905$) and $p < .000$. The ANOVA table revealed that Model 3 was significant at $F = 87.545$ ($df_3 = 2904$) and $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of variance in the LANGSCSC02 mean score from Model 1 to Model 2 and from Model 2 to Model 3. In Model 1, the change in R

Square = .036, Model 2 the change in R Square = .035 and Model 3 the change in R Square = .012.

The Beta was computed for each model. The Beta is a standardized score between -1 and +1 that identifies the strength of the relationship between variables with +1 indicating the strongest relationship. In Model 1, the standardized coefficient for the predictor of LANGSCSC02 was significant for WSRMODEL at $\beta = -.189$, $t = -10.378$, $p < .000$. In Model 2, the standardized coefficients for the predictors of LANGSCSC02 was significant for WSRMODEL at $\beta = -.191$, $t = -10.077$, $p < .000$ and was significant for GENDER02 at $\beta = .187$, $t = 10.441$, $p < .000$. In Model 3, the standardized coefficients for the predictors of LANGSCSC02 was significant for WSRMODEL at $\beta = -.168$, $t = -9.723$, $p < .000$, was significant for GENDER02 at $\beta = .187$, $t = 10.517$, $p < .000$ and was significant for ETHNICITY02 at $\beta = .113$, $t = 6.247$, $p < .000$.

Table 35.

Regression For Language Arts Literacy For 2002

Variable	B	SE B	β	R ²
Step 1				
WSRMODEL	-8.380	.807	-.189*	.036
Step 2				
WSRMODEL	-8.466	.793	-.191*	
GENDER02	8.006	.767	.187*	.071
Step 3				
WSRMODEL	-7.445	.805	-.168*	

GENDER02	8.012	.762	.187*	
ETHNICITY02	2.297	.368	.113*	.093

Multiple regressions were conducted to determine the effects of gender (GENDER03), ethnicity (ETHNICITY03) and the WSR Model (WSRMODEL) on Language Arts Literacy scores (LANGSCSC03). The dependent variable was LANGSCSC03 and the independent variables were GENDER03, ETHNICITY03 and WSRMODEL. The results of the analysis can be found in Table 36.

The Regression analysis yielded a LANGSCSC03 mean of 205.64. The correlations revealed a significant Pearson $r = -.244$ between LANGSCSC03 and WSRMODEL, a significant Pearson $r = .174$ between LANGSCSC03 and GENDER03, and a significant Pearson $r = .235$ between LANGSCSC03 and ETHNICITY03. Three models were constructed and in Model 1 R Square = .059, Model 2 R Square = .090 and Model 3 R Square = .130. The ANOVA revealed that Model 1 was significant at $F = 163.874$ ($df\ 1 = 2597$) and $p < .000$. The ANOVA revealed that Model 2 was significant at $F = 127.881$ ($df\ 2 = 2596$) and $p < .000$. The ANOVA revealed that Model 3 was significant at $F = 128.753$ ($df\ 3 = 2595$) and $p < .000$. In Model 1, the standardized coefficient for the predictor of LANGSCSC03 was significant for WSRMODEL at $\beta = -.244$, $t = -12.801$, $p < .000$. In Model 2, the standardized coefficients for the predictors of LANGSCSC03 was significant for WSRMODEL at $\beta = -.244$, $t = -13.011$, $p < .000$ and was significant for GENDER03 at $\beta = .174$, $t = 9.300$, $p < .000$. In Model 3, the standardized coefficients for the predictors of LANGSCSC03 was significant for WSRMODEL at $\beta = -.209$, $t = -11.264$, $p < .000$, was significant for GENDER03 at $\beta =$

.178, $t = 9.698$, $p < .000$ and was significant for ETHNICITY03 at $\beta = .203$, $t = 10.903$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the LANGSCSC03 mean score from Model 1 to Model 2 and from Model 2 to Model 3. In Model 1, the change in R Square = .059, Model 2 the change in R Square = .030 and Model 3 the change in R Square = .040.

Table 36.

Regression For Language Arts Literacy For 2003

Variable	B	SE B	β	R ²
Step 1				
WSRMODEL	-11.159	.872	-.244*	.059
Step 2				
WSRMODEL	-11.160	.858	-.244*	
GENDER03	7.925	.852	.174*	.090
Step 3				
WSRMODEL	-9.586	.851	-.209*	
GENDER03	8.084	.834	.178*	
ETHNICITY03	5.406	.496	.203*	.130

Multiple regressions were conducted to determine the effects of gender (GENDER04), ethnicity (ETHNICITY04) and the WSR Model (WSRMODEL) on Language Arts Literacy mean scores (LANGSCSC04). The dependent variable was

LANGSCSC04 and the independent variables were GENDER04, ETHNICITY04 and WSRMODEL. The results of the analysis can be found in Table 37.

The Regression analysis revealed a LANGSCSC04 mean of 208.60. The correlations yielded a significant Pearson $r = -.217$ between LANGSCSC04 and WSRMODEL, a significant Pearson $r = .112$ between LANGSCSC04 and GENDER04, and a significant Pearson $r = .211$ between LANGSCSC04 and ETHNICITY04. Three models were constructed and in Model 1 R Square = .047, Model 2 R Square = .058 and Model 3 R Square = .087. The ANOVA revealed that Model 1 was significant at $F = 114.143$ ($df\ 1 = 2309$) and $p < .000$. The ANOVA revealed that Model 2 was significant at $F = 71.5019$ ($df\ 2 = 2308$) and $p < .000$. The ANOVA revealed that Model 3 was significant at $F = 72.828$ ($df\ 3 = 2307$) and $p < .000$. In Model 1, the standardized coefficient for the predictor of LANGSCSC04 was significant for WSRMODEL at $\beta = -.217$, $t = -10.684$, $p < .000$. In Model 2, the standardized coefficients for the predictors of LANGSCSC04 was significant for WSRMODEL at $\beta = -.214$, $t = -10.600$, $p < .000$ and was significant for GENDER04 at $\beta = .106$, $t = 5.249$, $p < .000$. In Model 3, the standardized coefficients for the predictors of LANGSCSC04 was significant for WSRMODEL at $\beta = -.177$, $t = -8.663$, $p < .000$, was significant for GENDER04 at $\beta = .107$, $t = 5.366$, $p < .000$ and was significant for ETHNICITY04 at $\beta = .172$, $t = 8.434$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the LANGSCSC04 mean score from the Model 1 to Model 2 and from Model 2 to Model 3. Model 1 the change in R

Square = .047, Model 2 the change in R Square = .011 and Model 3 the change in R Square = .028.

Table 37.

Regression For Language Arts Literacy For 2004

Variable	B	SE B	β	R ²
Step 1				
WSRMODEL	-10.847	1.015	-.217*	.047
Step 2				
WSRMODEL	-10.704	1.010	-.214*	
GENDER04	5.131	.978	.106*	.058
Step 3				
WSRMODEL	-8.831	1.019	-.177*	
GENDER04	5.168	.963	.107*	
ETHNICITY04	4.901	.581	.172*	.087

This research question was concerned with the effects of gender, ethnicity and the WSR model on Language Arts Literacy. The analysis attempted to determine if the regression models helped predict mean scores and what variable had the greatest impact on mean scores. The analysis for this question indicated that all three models for 2002 were significant and all of the variables in all three models were significant. In Model 1, WSRMODEL was significant. In Model 2, WSRMODEL was the strongest predictor of mean score and, in Model 3, GENDER02 was the strongest predictor of mean score followed by WSRMODEL. All 3 models for 2003 were significant and all of the

variables in all three models were significant. In Model 2, WSRMODEL was the strongest predictor of mean score and, in Model 3, WSRMODEL was the strongest predictor of mean score followed by ETHNICITY03. All three models for 2004 were significant and all of the variables in all three models were significant. In Model 2, WSRMODEL was the strongest predictor of mean score followed by GENDER04.

The analyses revealed that each of the models during each year of this study were significant. In Model 1 from 2002-2004, the ability of WSRMODEL to predict Language Arts Literacy mean scores was significant each year. In Model 2, during the last two years of the study, WSRMODEL was the strongest predictor of LANGSCSC03 and LANGSCSC04. The analyses also revealed that, in Model 3 during the last two years of the study, WSRMODEL was the strongest predictor of Language Arts Literacy mean score LANGSCSC03 and LANGSCSC04. These findings provide more information to support the premise that WSR has an effect of Language Arts Literacy mean scores. In fact, regarding all of the variables used, it was the strongest predictor of Language Arts Literacy mean scores. Therefore, the effects of WSR can help to predict student achievement outcomes as measured by Language Arts Literacy mean scores.

Research Question 10

Are gender, ethnicity and the WSR model good predictors of Reading Cluster mean scores for school years 2002, 2003 and 2004?

Multiple regressions were conducted to determine the effects of gender (GENDER02), ethnicity (ETHNICITY02) and the WSR Model (WSRMODEL) on Reading Cluster mean scores (READING02). The dependent variable was READING02

and the independent variables were GENDER02, ETHNICITY02 and WSRMODEL. The results of the analysis can be found in Table 38.

The Regression analysis revealed a READING02 mean of 10.000. The correlations revealed a significant Pearson $r = -.148$ between READING02 and WSRMODEL, a significant Pearson $r = .154$ between READING02 and GENDER02, and a significant Pearson $r = .120$ between READING02 and ETHNICITY02. The R Square was computed for each model. The R Square identified the proportion of the variance in the LANGSCSC02 mean score that can be attributed to the model. In Model 1 R Square = .022, Model 2 R Square = .046 and Model 3 R Square = .055. The ANOVA revealed that Model 1 was significant at $F = 65.388$ ($df\ 1 = 2906$) and $p < .000$. The ANOVA revealed that Model 2 was significant at $F = 70.338$ ($df\ 2 = 2905$) and $p < .000$. The ANOVA revealed that Model 3 was significant at $F = 56.048$ ($df\ 3 = 2904$) and $p < .000$.

The Beta was computed for each model. The Beta is a standardized score between -1 and +1 that identifies the strength of the relationship between variables with +1, indicating the strongest relationship. In Model 1, the standardized coefficient for the predictor of READING02 was significant for WSRMODEL at $\beta = -.148$, $t = -8.086$, $p < .000$. In Model 2, the standardized coefficients for the predictors of READING02 was significant for WSRMODEL at $\beta = -.150$, $t = -8.275$, $p < .000$ and was significant for GENDER02 at $\beta = .156$, $t = 8.582$, $p < .000$. In Model 3, the standardized coefficients for the predictors of READING02 was significant for WSRMODEL at $\beta = -.131$, $t = -7.097$, $p < .000$, was significant for GENDER02 at $\beta = .156$, $t = 8.626$, $p < .000$ and was significant for ETHNICITY02 at $\beta = .094$, $t = 5.123$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING02 mean score from Model 1 to Model 2 and from Model 2 to Model 3. In Model 1 the change in R Square = .022, Model 2 the change in R Square = .024 and Model 3 the change in R Square = .009.

Table 38.

Regression For Reading Cluster For 2002

Variable	B	SE B	β	R ²
Step 1				
WSRMODEL	-1.110	.137	-.148*	.022
Step 2				
WSRMODEL	-1.122	.136	-.150*	
GENDER02	1.126	.131	.156*	.046
Step 3				
WSRMODEL	-.979	.138	-.131*	
GENDER02	1.127	.131	.156*	
ETHNICITY02	.323	.063	.094*	.055

Multiple regressions were conducted to determine the effects of gender (GENDER03), ethnicity (ETHNICITY03) and the WSR Model (WSRMODEL) on Reading Cluster mean scores (READING03). The dependent variable is READING03 and the independent variables are GENDER03, ETHNICITY03 and WSRMODEL. The results of the analysis can be found in Table 39.

The Regression analysis yielded a READING03 mean of 9.245. The correlations revealed a significant Pearson $r = -.224$ between READING03 and WSRMODEL, a significant Pearson $r = .131$ between READING03 and GENDER03, and a significant Pearson $r = .216$ between READING03 and ETHNICITY03. Three models were constructed and in Model 1 R Square = .050, Model 2 R Square = .067 and Model 3 R Square = .101. The ANOVA table revealed that Model 1 was significant at $F = 137.409$ ($df\ 1 = 2605$) and $p < .000$. The ANOVA table revealed that Model 2 was significant at $F = 93.744$ ($df\ 2 = 2604$) and $p < .000$. The ANOVA table revealed that Model 3 was significant at $F = 97.049$ ($df\ 3 = 2603$) and $p < .000$. In Model 1, the standardized coefficient for the predictor of READING03 was significant for WSRMODEL at $\beta = -.224$, $t = -11.722$, $p < .000$. In Model 2, the standardized coefficients for the predictors of READING03 was significant for WSRMODEL at $\beta = -.224$, $t = -11.821$, $p < .000$ and was significant for GENDER03 at $\beta = .131$, $t = 6.901$, $p < .000$. In Model 3, the standardized coefficients for the predictors of READING03 was significant for WSRMODEL at $\beta = -.192$, $t = -10.186$, $p < .000$, was significant for GENDER03 at $\beta = .134$, $t = 7.194$, $p < .000$ and was significant for ETHNICITY03 at $\beta = .186$, $t = 9.837$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING03 mean score from Model 1 to Model 2 and from Model 2 to Model 3. In Model 1 the change in R Square = .050, Model 2 the change in R Square = .017 and Model 3 the change in R Square = .033.

Table 39.

Regression For Reading Cluster For 2003

Variable	B	SE B	β	R ²
Step 1				
WSRMODEL	-1.954	.167	-.224*	.050
Step 2				
WSRMODEL	-1.953	.165	-.224*	
GENDER03	1.133	.164	.131*	.067
Step 3				
WSRMODEL	-1.677	.165	-.192*	
GENDER03	.1.160	.161	.134*	
ETHNICITY03	.944	.098	.186*	.101

Multiple regressions were conducted to determine the effects of gender (GENDER04), ethnicity (ETHNICITY04) and the WSR Model (WSRMODEL) on Reading Cluster mean scores (READING04). The dependent variable is READING04 and the independent variables are GENDER04, ETHNICITY04 and WSRMODEL. The results of the analysis can be found in Table 40.

The Regression analysis yielded a READING04 mean of 11.39. The correlations revealed a significant Pearson $r = -.189$ between READING04 and WSRMODEL, a significant Pearson $r = .073$ between READING04 and GENDER04, and a significant Pearson $r = .213$ between READING04 and ETHNICITY04. Three models were constructed and in Model 1 R Square = .036, Model 2 R Square = .041 and Model 3 R

Square = .071. The ANOVA table revealed that Model 1 was significant at $F = 85.953$ (df 1 = 2309) and $p < .000$. The ANOVA table revealed that Model 2 was significant at $F = 48.710$ (df 2 = 2308) and $p < .000$. The ANOVA table revealed that Model 3 was significant at $F = 59.129$ (df 3 = 2307) and $p < .000$. In Model 1, the standardized coefficient for the predictor of READING04 was significant for WSRMODEL at $\beta = -.189$, $t = -9.271$, $p < .000$. In Model 2, the standardized coefficients for the predictors of READING04 was significant for WSRMODEL at $\beta = -.188$, $t = -9.199$, $p < .000$ and was significant for GENDER04 at $\beta = .068$, $t = 3.330$, $p < .001$. In Model 3, the standardized coefficients for the predictors of READING04 was significant for WSRMODEL at $\beta = -.148$, $t = -7.214$, $p < .000$, was significant for GENDER04 at $\beta = .069$, $t = 3.425$, $p < .001$ and was significant for ETHNICITY04 at $\beta = .180$, $t = 8.762$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING04 mean score from Model 1 to Model 2 and from Model 2 to Model 3. In Model 1 the change in R Square = .036, Model 2 the change in R Square = .005 and Model 3 the change in R Square = .031.

Table 40.

Regression For Reading Cluster For 2004

Variable	B	SE B	β	R^2
Step 1				
WSRMODEL	-1.622	.175	-.189*	.036
Step 2				

WSRMODEL	-1.606	.175	-.188*	
GENDER04	.563	.169	.068*	.041
Step 3				
WSRMODEL	-1.270	.176	-.148*	
GENDER04	.570	.166	.069*	
ETHNICITY04	.879	.100	.180*	.071

This research question was concerned with the effect of gender, ethnicity and the WSR model on Reading Cluster. The analysis attempted to determine if the regression models can help to predict mean scores and what variable had the greatest impact on mean scores. The analysis for this question indicated that all three models for 2002 were significant and all of the variables in all three models were significant. In Model 1, WSRMODEL was significant. In Model 2, GENDER02 was the strongest predictor of mean score and, in Model 3, GENDER02 was the strongest predictor of mean score followed by WSRMODEL. All three models for 2003 were significant and all of the variables in all three models were significant. In Model 2, WSRMODEL was the strongest predictor of mean score and, in Model 3, WSRMODEL was the strongest predictor of mean score followed by ETHNICITY03. All three models for 2004 were significant and all of the variables in all three models were significant. In Model 2, WSRMODEL was the strongest predictor of mean score and, in Model 3, ETHNICITY04 was the strongest predictor of mean score followed by WSRMODEL.

The analyses revealed that each of the models during each year of this study were significant. In Model 1 from 2002-2004, the ability of WSRMODEL to predict Reading Cluster mean scores was significant each year. It also revealed that, in Model 3, during

the last two years of the study, WSRMODEL was the strongest predictor of READING04 mean scores. These findings provide more information to support the premise that WSR has an effect on Reading Cluster mean scores. In fact, regarding all of the variables it was the strongest predictor of Reading Cluster mean scores. Therefore, the effects of WSR can help to predict student achievement outcomes as measured by Reading Cluster mean scores.

Research Question 11

Are gender, ethnicity and the WSR model good predictors of Non-SFA schools Reading Cluster mean scores for school years 2002, 2003 and 2004?

Multiple regressions were conducted to determine the effects of gender (GENDER02) and ethnicity (ETHNICITY02) on Non-SFA Reading Cluster mean scores (READING02). The dependent variable was READING02 and the independent variables were GENDER02 and ETHNICITY02. SFA schools were coded 2 and Non-SFA schools were coded 1. The results of the analysis can be found in Table 41.

The Regression analysis revealed a READING02 mean of 9.309. The correlations revealed a significant Pearson $r = .182$ between READING02 and GENDER02, and a non-significant Pearson $r = .021$ between READING02 and ETHNICITY02. The R Square was computed for each model. The R Square identifies the proportion of the variance in the READING02 mean score that can be attributed to the model. In Model 1 R Square = .033. The ANOVA revealed the Model 1 was significant at $F = 37.596$ ($df\ 1 = 1096$) and $p < .000$.

The Beta was computed for each model. The Beta is a standardized score between -1 and +1 that identifies the strength of the relationship between variables, with +1

indicating the strongest relationship. In Model 1, the standardized coefficient for the predictor of READING02 was significant for GENDER02 at $\beta = .182$, $t = 6.132$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING02 mean score from Model 1 to Model 2. In Model 1, the change in R Square = .033.

Table 41.

Regression For Non-SFA Reading Cluster For 2002

Variable	B	SE B	β	R ²
Step 1				
GENDER02	7.441	.321	.182*	.033

Multiple regressions were conducted to determine the effects of gender (GENDER03) and ethnicity (ETHNICITY03) on Non-SFA Reading Cluster mean scores (READING03). The dependent variable was READING03 and the independent variables were GENDER03 and ETHNICITY03. SFA schools were coded 2 and Non-SFA schools were coded 1. The results of the analysis can be found in Table 42.

The Regression analysis revealed a READING03 mean of 8.165. The correlations revealed a significant Pearson $r = .094$ between READING03 and GENDER03, and a significant Pearson $r = .194$ between READING03 AND ETHNICITY03. Two models were constructed and in Model 1 R Square = .038 and in Model 2 R Square = .042. The ANOVA revealed that Model 1 was significant at $F = 45.591$ ($df 1 = 1163$) and $p < .000$. The ANOVA revealed that Model 2 was significant at $F = 25.205$ ($df 2 = 1162$) and $p <$

.000. In Model 1, the standardized coefficient for the predictor of READING03 was significant for GENDER03 at $\beta = .194$, $t = 6.752$, $p < .000$. In Model 2, the standardized coefficient for the predictors of READING03 was significant for GENDER03 at $\beta = .193$, $T = 3.629$, $p < .000$ and for ETHNICITY03 at $\beta = .062$, $t = 2.162$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING03 mean score from Model 1 to Model 2. In Model 1 the change in R Square = .038 and Model 2 the change in R Square = .004.

Table 42.

Regression For Non-SFA Reading Cluster For 2003

Variable	B	SE B	β	R^2
Step 1				
GENDER03	.779	.115	.194*	.038
Step 2				
GENDER03	.776	.115	.193*	
ETHNICITY03	.538	.249	.062*	.042

Multiple regressions were conducted to determine the effects of gender (GENDER04) and ethnicity (ETHNICITY04) on Non-SFA schools Reading Cluster mean scores (READING04). The dependent variable was READING04 and the independent variables were GENDER04 and ETHNICITY04. SFA schools were coded 2 and Non-SFA schools were coded 1. The results of the analysis can be found in Table 43.

The Regression analysis revealed a READING04 mean of 10.39. The correlations revealed a significant Pearson $r = .123$ between READING04 and GENDER04, and a non-significant Pearson $r = .030$ between READING04 and ETHNICITY04. Two models were constructed and in Model 1 R Square = .015. The ANOVA revealed that Model 1 was significant at $F = 13.693$ ($df 1 = 893$) and $p < .000$. In Model 1, the standardized coefficient for the predictor of READING04 was significant for ETHNICITY04 at $\beta = .123$, $t = 3.700$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING04 mean score from Model 1 to Model 2. In Model 1 the change in R Square = .015.

Table 43.

Regression For Non-SFA Reading Cluster For 2004

Variable	B	SE B	β	R ²
Step 1				
GENDER04	1.020	.276	.123*	.015

This research question was concerned with the effects of gender and ethnicity on Non-SFA schools Reading Cluster means. The analyses attempted to determine if the regression models can help predict mean scores and what variable had the greatest impact on mean scores. The analysis for this question indicated that for 2002, one model was significant and GENDER02 was the only significant variable. In 2003, all of the models were significant and all of the variables in both models were significant. In Model 2,

GENDER03 was the strongest predictor of mean score. In 2004, one model was significant and GENDER04 was the only significant variable.

The analyses revealed that four out of six models were significant and gender was significant in each of the four significant models. It also revealed that, in Model 2, where gender and ethnicity were significant predictors of Reading Cluster mean scores, gender was the greatest predictor. These findings provide more information to support the premise that gender can affect Non-SFA Reading Cluster mean scores. In fact, regarding all of the variables, it was the strongest predictor of Reading Cluster mean scores. Therefore, the effects of gender can help to predict student achievement outcomes as measured by Non-SFA schools Reading Cluster mean scores.

Research Question 12

Are gender, ethnicity and the WSR model good predictors of SFA schools Reading Cluster mean scores for school years 2002, 2003 and 2004?

Multiple regressions were conducted to determine the effects of gender (GENDER02) and ethnicity (ETHNICITY02) on SFA schools Reading Cluster mean scores (READING02). The dependent variable was READING02 and the independent variables were GENDER02 and ETHNICITY02. SFA schools were coded 2 and Non-SFA schools were coded 1. The results of the analysis can be found in Table 44.

The Regression analysis revealed a READING02 mean of 10.420. The correlations revealed a significant Pearson $r = .144$ between READING02 and GENDER02, and a significant Pearson $r = .170$ between READING02 and ETHNICITY02. The R Square was computed for each model. The R Square identifies the proportion of the variance in the READING02 mean score that can be attributed to

the model. In Model 1 R Square = .021 and in Model 2 R Square = .029. The ANOVA revealed that Model 1 was significant at $F = 38.302$ ($df 1 = 1808$) and $p < .000$. The ANOVA revealed that Model 2 was significant at $F = 46.994$ ($df 2 = 1807$) and $p < .000$.

The Beta was computed for each model. The Beta is a standardized score between -1 and +1 that identifies the strength of the relationship between variables, with +1 indicating the strongest relationship. In Model 1, the standardized coefficient for the predictor of READING02 was significant for GENDER02 at $\beta = .144$, $t = 6.189$, $p < .000$. In Model 2, the standardized coefficient for the predictor of READING02 was significant for GENDER02 at $\beta = .143$, $t = 6.234$, $p < .000$ and was significant for ETHNICITY02 at $\beta = .169$, $t = 7.386$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING02 mean score from Model 1 to Model 2. In Model 1 the change in R Square = .021 and Model 2 the change in R Square = .029.

Table 44.

Regression For SFA Reading Cluster For 2002

Variable	B	SE B	β	R ²
Step 1				
GENDER02	1.068	.173	.144*	.021
Step 2				
GENDER02	1.060	.170	.143*	.021
ETHNICITY02	1.263	.171	.169*	.029

Multiple regressions were conducted to determine the effects of gender (GENDER03) and ethnicity (ETHNICITY03) on SFA Reading Cluster mean scores (READING03). The dependent variable was READING03 and the independent variables were GENDER03 and ETHNICITY03. SFA schools were coded 2 and Non-SFA schools were coded 1. The results of the analysis can be found in Table 45.

The Regression analysis revealed a READING03 mean of 10.119. The correlations revealed a significant Pearson $r = .091$ between READING03 and GENDER03, and a significant Pearson $r = .099$ between READING03 and ETHNICITY03. Two models were constructed and in Model 1 R Square = .008 and in Model 2 R Square = .019. The ANOVA revealed that Model 1 was significant at $F = 11.922$ ($df 1 = 1440$) and $p < .000$. The ANOVA revealed that Model 2 was significant at $F = 13.947$ ($df 2 = 1439$) and $p < .000$. In Model 1, the standardized coefficient for the predictor of READING03 was significant for GENDER03 at $\beta = .091$, $t = 3.453$, $p < .000$. In Model 2, the standardized coefficient for the predictors of READING03 were significant for GENDER03 at $\beta = .096$, $t = 3.672$, $p < .000$ and for ETHNICITY03 at $\beta = .104$, $t = 3.981$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING03 mean score from Model 1 to Model 2. In Model 1 the change in R Square = .008 and Model 2 the change in R Square = .011.

Table 45.

Regression For SFA Reading Cluster For 2003

Variable	B	SE B	β	R ²
Step 1				
GENDER03	.399	.116	.091*	.008
Step 2				
GENDER03	.423	.115	.096*	
ETHNICITY03	.733	.184	.104*	.019

Multiple regressions were conducted to determine the effects of gender (GENDER04) and ethnicity (ETHNICITY04) on SFA schools Reading Cluster mean scores (READING04). The dependent variable was READING04 and the independent variables were GENDER04 and ETHNICITY04. SFA schools were coded 2 and Non-SFA schools were coded 1. The results of the analysis can be found in Table 46.

The Regression analysis revealed a READING04 mean of 12.02. The correlations revealed a non-significant Pearson $r = .037$ between READING04 and GENDER04 and a significant Pearson $r = .164$ between READING04 and ETHNICITY04. Two models were constructed and in Model 1 R Square = .027. The ANOVA revealed that Model 1 was significant at $F = 39.091$ ($df 1 = 1414$) and $p < .000$. In Model 1, the standardized coefficient for the predictor of READING04 was significant for ETHNICITY04 at $\beta = .164$, $t = 6.252$, $p < .000$.

The change in R Square was computed for each model. The change in R Square identifies the change of the proportion of the variance in the READING04 mean score from Model 1 to Model 2. In Model 1 the change in R Square = .027.

Table 46.

Regression For SFA Reading Cluster For 2004

Variable	B	SE B	β	R ²
Step 1				
ETHNICITY04	1.076	.172	.164*	.027

This research question was concerned with the effects of gender and ethnicity on SFA schools Reading Cluster means. The analyses attempted to determine if the regression models could help predict mean scores and what variable had the greatest impact on mean scores. The analysis for this question indicated that for 2002, all of the models were significant and all of the variables in both models were significant. In Model 2, ETHNICITY02 was the strongest predictor of mean score. In 2003, all of the models were significant and all of the variables in both models were significant. In Model 2, ETHNICITY03 was the strongest predictor of mean score. In 2004, one model was significant and ETHNICITY04 was the only significant variable.

The analyses revealed that five out of six models were significant. Gender was significant in four out of six models, where it was entered, and ethnicity is significant in each of the three significant models, where it is entered. It also revealed that, in Model 2, where gender and ethnicity were significant predictors of Reading Cluster mean scores, ethnicity was the greatest predictor. These findings provide more information to support

the premise that gender and ethnicity can have an effect on SFA Reading Cluster mean scores. In fact, when both variables were significant in a model, ethnicity was the strongest predictor of Reading Cluster mean scores. Therefore, the effects of gender and ethnicity can help to predict student achievement outcomes as measured by SFA schools Reading Cluster mean scores.

CHAPTER V

SUMMARY AND CONCLUSIONS AND RECOMMENDATIONS

This chapter is divided into four sections; Summary, Conclusions, Policy Recommendations and Future Research.

Summary

The purpose of this study was to investigate the impact of the SFA whole-school reform model on student achievement, as measured by the ESPA and NJ ASK4 in the Urban Public Schools. The following research questions were explored during this study:

1. Are there significant differences among the cohorts when comparing Language Arts Literacy mean scores from 2002-2004?
2. Are there significant differences among the cohorts when comparing Reading Cluster mean scores from 2002-2004?
3. Are there significant differences between School Group A and other School Groups when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
4. Are there significant differences between School Group A and other School Groups when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?
5. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Language Arts Literacy mean scores from 2002-2003, 2003-2004 and 2002-2004?

6. Are there significant differences within the SFA schools and within the Non-SFA schools when comparing Reading Cluster mean scores from 2002-2003, 2003-2004 and 2002-2004?
7. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
8. Are there significant differences between the group of SFA schools and the group of Non-SFA schools when comparing Reading Cluster mean scores for school years 2002, 2003 and 2004?
9. Are gender, ethnicity and the WSR model good predictors of Language Arts Literacy mean scores for school years 2002, 2003 and 2004?
10. Are gender, ethnicity and the WSR model good predictors of Reading Cluster mean scores for school years 2002, 2003 and 2004?
11. Are gender and ethnicity good predictors of Non-SFA schools Reading Cluster mean scores for school years 2002, 2003 and 2004?
12. Are gender and ethnicity good predictors of SFA schools Reading Cluster mean scores for school years 2002, 2003 and 2004?

The necessary data was obtained from the Urban Public Schools and the New Jersey Department of Education. T-Tests and regression analyses were conducted using SPSS to obtain answers to the research questions.

The first research question was concerned with significant differences on a district level when comparing Language Arts Literacy mean scores from 2002 to 2004. The analysis attempted to determine if WSR is having an impact on student achievement.

“Based on research and experience, advocates of comprehensive school-wide reform believe that it holds the potential for raising student achievement not only for disadvantaged students but for all students” (Protheroe, 1998, p. 7). The analysis for this question resulted in a significant decrease in the district Language Arts Literacy mean score from 2002 to 2003 and a significant increase in the district Language Arts Literacy mean score from 2003 to 2004. However, the increase in Language Arts Literacy mean score from 2002 to 2004 was not significant. This is consistent with Pogrow’s findings that, “Neither the general school-wide approach nor the extant school-wide designs have demonstrated substantial effectiveness” (Pogrow, p. 74).

The second research question was concerned with significant differences on a district level when comparing Reading Cluster mean scores from 2002 to 2004. The analysis attempted to determine if WSR is having an impact on student achievement. The analysis for this question resulted in a significant decrease in the district Reading Cluster mean score from 2002 to 2003 and a significant increase in the district Reading Cluster mean score from 2003 to 2004. In addition, the increase in the Reading Cluster mean score from 2002 to 2004 was significant. According to Singer (2002), “the improvement of growth in scores has been positive in LAL for all WSR models and, . . .when all schools are studied together without regard for their specific WSR Models, the growth is also shown to be significant” (p. 135).

The third research question was concerned with significant Language Arts Literacy differences between School Group A and the combination of the other School Groups. The analysis attempted to determine the impact of the majority presence of the SFA schools in School Group A on student achievement. The analysis for this question

resulted in a significant difference in Language Arts Literacy mean scores between School Group A and other School Groups favoring the other School Groups in 2002, 2003 and 2004.

The fourth research question was concerned with significant Reading Cluster differences between School Group A and the combination of the other School Groups. The analysis attempted to determine the impact of the majority presence of the SFA schools in School Group A on student achievement. The analysis for this question resulted in a significant difference in Reading Cluster mean scores between School Group A and other School Groups, favoring the other School Groups in 2002, 2003 and 2004.

The fifth research question was concerned with significant differences on a WSR level when comparing Language Arts Literacy mean scores from 2002 to 2004. The analysis attempted to determine if WSR had an impact on student achievement. The analysis for this question resulted in significant Language Arts Literacy increases for Non-SFA schools from 2003-2004 and 2002-2004. Also, the analysis for this question resulted in a significant Language Arts Literacy increase for SFA schools from 2003-2004. The analysis for this question resulted in a significant Language Arts Literacy decrease for SFA schools from 2002-2003.

The sixth research question was concerned with significant differences on a WSR level when comparing Reading Cluster mean scores from 2002 to 2004. The analysis attempted to determine if WSR had an impact on student achievement. The analysis for this question resulted in significant Reading Cluster increases for Non-SFA schools from 2003-2004 and 2002-2004. The analysis for this question also resulted in a significant decrease for Non-SFA schools from 2002-2003. The analysis for this question resulted in

significant Reading Cluster increases for SFA schools from 2003-2004 and 2002-2004. The analysis for this question also resulted in a significant Reading Cluster decrease for SFA schools from 2002-2003.

The seventh research question was concerned with significant Language Arts Literacy differences between the group of Non-SFA schools and the group of SFA schools. The analysis attempted to determine the Language Arts Literacy differences between the group of Non-SFA schools and the group of SFA schools. The analysis for this question resulted in significant differences in Language Arts Literacy mean scores between the group of Non-SFA schools and the group of SFA schools in 2002, 2003 and 2004.

The eighth research question was concerned with significant Reading Cluster differences between the group of Non-SFA schools and the group of SFA schools. The analyses attempted to determine the Reading Cluster mean differences between the group of Non-SFA schools and the group of SFA schools. The analysis for this question resulted in significant differences in Reading Cluster mean scores between the group of Non-SFA schools and the group of SFA schools, favoring the Non-SFA schools in 2002 and 2003.

The ninth research question was concerned with the effects of gender, ethnicity and the WSR model on Language Arts Literacy. The analysis attempted to determine if the regression models helped predict mean scores and what variable had the greatest impact on mean scores. The analysis for this question indicated that all three models for 2002 were significant and all of the variables in all three models were significant. In Model 2, WSR was the greatest predictor of mean score and, in Model 3, gender was the

strongest predictor of mean score followed by WSR. All three models for 2003 were significant and all of the variables in all three models were significant. In Model 2, WSR was the strongest predictor of mean score and, in Model 3, WSR was the strongest predictor of mean score followed by ethnicity. All three models for 2004 were significant and all of the variables in all three models were significant. In Model 2, WSR was the strongest predictor of mean score and, in Model 3, WSR was the strongest predictor of mean score followed by gender.

The tenth research question was concerned with the effects of gender, ethnicity and the WSR model on Reading Cluster mean scores. The analysis attempted to determine if the regression models helped predict mean scores and what variables had the greatest impact on mean scores. The analysis for this question indicated that all three models for 2002 were significant and all of the variables in all three models were significant. In Model 2, gender was the strongest predictor of mean score and in Model 3 gender was the strongest predictor of mean score followed by WSR. All three models for 2003 were significant and all of the variables in all three models were significant. In Model 2, WSR was the strongest predictor of mean score and in Model 3 WSR was the strongest predictor of mean score, followed by ethnicity. All three models for 2004 were significant and all of the variables in all three models were significant. In Model 2, WSR was the strongest predictor of mean score and in Model 3 ethnicity was the strongest predictor of mean score, followed by WSR.

The eleventh research question was concerned with the effects of gender and ethnicity on Non-SFA schools Reading Cluster mean scores. The analysis attempted to determine if the regression models helped predict mean scores and what variables had the

greatest impact on mean scores. The analysis for this question indicated that, for 2002, one model was significant and gender was the only significant variable. In 2003, all of the models were significant and all of the variables in both models were significant. In Model 2, gender was the strongest predictor of mean score. In 2004, one model was significant and gender was the only significant variable.

The twelfth research question was concerned with the effects of gender and ethnicity on SFA schools Reading Cluster means scores. The analysis attempted to determine if the regression models helped predict mean scores and what variables had the greatest impact on mean scores. The analysis for this question indicated that for 2002, all of the models were significant and all of the variables in both models were significant. In Model 2, ethnicity was the strongest predictor of mean score. In 2003, all of the models were significant and all of the variables in both models were significant. In Model 2, ethnicity was the strongest predictor of mean score. In 2004, one model was significant and ethnicity was the only significant variable.

Conclusions

The results of this study revealed that District Language Arts Literacy and District Reading Cluster mean scores were the highest during the last year of this study and that the mean scores' increases from 2003 to 2004 were significant. Language Arts Literacy mean scores for Non-SFA schools were the highest during the last year of this study and the increases from 2003 to 2004 were significant. In addition, the Non-SFA schools Language Arts Literacy mean score increases were significant from 2002 to 2004. The Reading Cluster mean scores for Non-SFA schools and SFA schools were the highest

during the last year of this study and the increases from 2003 to 2004 and from 2002 to 2004 were significant.

The results of this study also revealed that Language Arts Literacy and Reading Cluster mean scores of School Group A were significantly lower than the mean scores of the other School Groups during each year of this study. Also, the Language Arts Literacy mean scores of the SFA schools were significantly lower than the mean scores of the Non-SFA schools during each year of this study. Reading Cluster mean scores for SFA schools were significantly lower than the mean scores of Non-SFA schools during each year of this study.

The results of this study revealed that, for 2002, gender was the best predictor of District Language Arts Literacy mean scores. Also, for 2003 and 2004, WSR was the best predictor of District Language Arts Literacy mean scores. In addition, for 2002, gender was the best predictor of District Reading Cluster mean scores and, for 2003 and 2004, WSR was the best predictor District Reading Cluster mean scores. In 2002, 2003 and 2004, gender was the best predictor of Non-SFA schools Reading Cluster mean scores; however, in 2002, 2003 and 2004, ethnicity was the best predictor of SFA schools Reading Cluster mean scores.

The results of this study revealed that WSR had a positive effect on ESPA and NJ ASK4 Language Arts Literacy and Reading Cluster mean scores. However, Non-SFA schools performed better on the ESPA and NJ ASK4 during the three years evaluated during this study. Therefore, the impact that SFA had on ESPA and NJ ASK4 has not been as great as the impact when comparing Non-SFA Language Arts Literacy and Reading Cluster mean scores. In addition, we must also remember that, in 2001, SFA was

the only WSR model with a state Language Arts Literacy mean score below the passing mean of 200. The State of New Jersey SFA Schools Language Arts Literacy mean scores were 176.95 for 1999, 177.75 for 2002 and 198.63 for 2001. The Urban Public Schools SFA schools Language Arts Literacy mean scores were 202.38 for 2002, 199.49 for 2003 and 201.96 for 2004.

Policy Recommendations

The SFA WSR model is no longer being implemented by the Urban Public Schools. The decision to stop using SFA was based on several reasons, with the primary reason being that, compared to other schools throughout the Urban Public Schools, SFA schools lagged behind Non-SFA schools regarding achieving NCLB benchmarks. However, as revealed by this study, the District, School Groups, Non-SFA and SFA schools consistently made Language Arts Literacy and Reading Cluster mean score increases. In addition, the Language Arts Literacy and Reading Cluster mean scores increased for the District, School Groups, Non-SFA and SFA schools during the last year of the study. Therefore, SFA and WSR should not be considered failures regarding improving student achievement. However, SFA can be considered a failure regarding achieving Language Arts Literacy and Reading Cluster mean scores that are significantly lower than Non-SFA school mean scores and not achieving NCLB benchmarks. The staff and administrators at the schools where SFA no longer exists must recognize this distinction in order for them to make effective curriculum decisions and to continue to use some of the effective instructional strategies that they employed while implementing SFA. According to McDonald (1999), new initiatives or reforms should not "...ignore the work of others who have gone before" (p. 3).

The New Jersey State Department of Education no longer considers SFA its presumptive model. The primary reason for this decision was based on the fact that SFA schools throughout the state were performing at a lower level than all other WSR models regarding mean scores and proficiency levels. The district and the state should also realize that SFA did not fail to improve student achievement; however, its failure was producing schools that achieved NCLB benchmarks and Language Arts Literacy and Reading Cluster mean scores at the level of Non-SFA school mean scores. The NJDOE must continue to support WSR in the Urban Public Schools because it has resulted in significant increases in academic achievement as measured by ESPA and NJ ASK4.

Future Research

This study can be replicated in the Urban Public Schools and in other districts. This study can also be replicated to evaluate the effectiveness of other former WSR models in the Urban Public Schools. In addition, this study can be replicated in other urban districts or regions using their respective state mandated assessments.

This study can be extended for the Urban Public Schools. This study can be extended to include Urban Public Schools NJ ASK4 data for 2004-2005 and 2005-2006 to determine if the positive trends are continuing. This study can also be extended by the appropriate number of years necessary to determine the impact of SFA on student achievement as measured by district and state standardized tests for students beyond grade 4, including the NJ ASK 5, NJ ASK 6, NJ ASK 7, GEPA and the HSPA.

This study can be expanded to evaluate the achievement of students that attended their respective SFA schools for the duration of the study. This study can also be expanded to evaluate the effects of SFA on Limited English Proficient and Special

Education Students. In addition, this study can be expanded to evaluate the effectiveness of SFA on a school level and to evaluate the effects of teacher fidelity, teaching proficiency and teaching experience regarding SFA on student achievement.

REFERENCES

- Abbott v. NJ, 710 A.2d 450,153 N.J. 480, 126 Ed. Law Rep. 258 (1998) (Supreme Court of New Jersey 1998).
- Anderson, G., & Arsenault, N. (1999). *Fundamentals of Educational Research* (2nd ed.). Philadelphia: The Falmer Press, Taylor & Francis Inc.
- Astuto, T. A., Clark, D. L., Read, A.-M., McGree, K., & Fernandez, L. d. P. (1994). *Roots of Reform: Challenging the Assumptions that Control Education Reform*. Bloomington: Phi Delta Kappa Educational Foundation.
- Babbie, E. (2002). *The Basics of Social Research* (Second Edition ed.). USA: Thompson Learning Inc.
- Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). *Facing the Challenges of Whole-School Reform: New American Schools After a Decade*: RAND Education.
- Berends, M., Chun, J., Schuyler, G., Stockly, S., & Briggs, R. J. (2002). *Challenges of Conflicting School Reforms: Effects of New American Schools in a High Poverty District*: RAND Education.
- Education, N. J. D. o. (2004a). *Abbott Implementation Whole School Reform In Abbott Districts Background Paper*. Retrieved November 20, 2004, 2004, from <http://www.nj.gov/njded/abbotts/archives/wsrback.htm>
- Education, N. J. D. o. (2004b). *Assessment and Evaluation*. Retrieved February 20, 2004, 2004, from <http://www.nj.gov/njded/assessment/>
- Education, N. J. D. O. (2004a). *Elementary School Statewide Assessments*. Retrieved May 11, 2004, 2004, from <http://www.nj.gov/njded/assessment/es/>
- Education, N. J. D. O. (2004b). *Historical Context: Overview of New Jersey's Statewide Testing Program*. Retrieved May 11, 2004, 2004, from <http://www.nj.gov/njded/assessment/history.shtml>

- Education, N. J. D. o. (2005). *DOE Announces NCLB-Designated Districts In Need of Improvement*. Retrieved March 28, 2005, 2005, from <http://www.state.nj.us/njded/news/2005/0307dini2.htm>
- Education, N. J. S. D. O. (2002). *May 2002 Elementary School Proficiency Assessment (ESPA) State Summary*.
- Education, U. S. D. o. (2004). *Introduction: No Child Left Behind*. Retrieved May 10, 2004, 2004, from <http://www.ed.gov/nclb/overview/intro/index.html>
- ETS. (2004). *New Jersey Department of Education Assessment of Skills and Knowledge (NJ ASK)*. Retrieved May 11, 2004, 2004, from <http://www.ets.org/njask/>
- George, D., & Mallery, P. (2001). *SPSS for Windows Step by Step: A Simple Guide and Reference 10.0 Update* (Third Edition ed.): Allyn & Bacon.
- Gordon, D. T. (Ed.). (2003). *A Nation Reformed? American Education 20 Years after A Nation At Risk*. Cambridge: Harvard Education Press.
- Guide for Implementing Urban Education Reform in Abbott Districts. 3.
- Gwynn, M. J., & Chase, J. B. J. (1969). *Curriculum Principles and Social Trends* (Fourth ed.). New York: The Macmillan Company.
- Heidenreich, R. R. (Ed.). (1972). *Current Readings In Improvements In Curriculum*. Arlington VA: College Readings Inc.
- Herman, R. (1999). *An Educators' Guide to Schoolwide Reform*. Annapolis: American Association of School Administrators.
- Herman, R., & Stringfield, S. (1997). *Ten Promising Programs for Educating All Children: Evidence of Impact*. Arlington: Educational Research Service.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (Fifth Edition ed.). Boston: Houghton Mifflin Company.

- Jones, E. M., Gottfredson, G. D., & Gottfredson, D. C. (1997). Success for Some: An Evaluation of a Success for All Program. *Evaluation Review*, v21(Number 6), 643-670.
- Keefe, J. W., Jenkins, J. M., & Hersey, P. W. (1992). *A Leaders Guide to School Restructuring: A Special Report of the NASSP Commission on Restructuring*. Reston: National Association of Secondary School Principals.
- Librera, W. L. (2003). *Designation of Abbott Districts Criteria and Process*. Retrieved April 24, 2004, from <http://www.nj.gov/njded/abbotts/regs/criteria.htm>
- Madden, N. A. (1993). Success for All: Longitudinal Effects of a Restructuring Program for Inner-City Elementary Schools. *American Educational Research Journal*, 30(Spring 1993), 123-148.
- Marsh, C. J., & Willis, G. (1999 1995). *Curriculum: Alternative Approaches, Ongoing Issues* (2nd ed.). Upper Saddle River: Prentice-Hall Inc.
- Marzano, R. J. (2003). *What Works In Schools: Translating Research Into Action*. Alexandria Virginia: Association for Supervision and Curriculum Development (ASCD).
- McDonald, J. P., Hatch, T., Kirby, E., Ames, N., Haynes, N. M., & Joyner, E. T. (1999). *School Reform Behind the Scenes How ATLAS is Shaping the Future of Education*. New York: Teachers College Press.
- Muirhead, M. S., Tyler, R. L., & Hamilton, M. P. (2001). *Study of Whole School Reform Implementation in New Jersey Abbott Districts: Perceptions of Key Stakeholders. Final Report*. Arlington: Region III Comprehensive Center.
- Overview of the School Development Program*. (2004). Retrieved April 19, 2004, from <http://www.info.med.yale.edu/comer/about/overview.html>
- Pogrow, S. *Title I: Wrong Help at the Wrong Time*. Retrieved August 13, 2002, from <http://www.edexcellence.net/library/newdrct.htm>
- Pogrow, S. (2000). Success for All Does Not Produce Success for Students. *Phi Delta Kappan*, September 2000.

- Protheroe, N. J. (1998). *Comprehensive Models for School Improvement: Finding the Right Match and Making It Work*. Arlington: Educational Research Service.
- Ravid, R. (2000). *Practical Statistics for Educators* (2nd ed.). Lanham: University Press of America, Inc.
- Rossi, R. J., & Stringfield, S. C. (1997). *Education Reform and Students at Risk*. Washington: U.S. Government Printing Office.
- Schools, N. A. (1998). *Blueprints for School Success: A Guide to New American Schools Designs*. Arlington: Educational Research Service.
- Schools, T. N. P. (2003). *Three Year Operational Plan 2003/04 - 2004/05*. Newark: The Newark Public Schools.
- Services, N. J. D. o. E. D. o. F. D. o. S. (2000). *Guide for Implementing Urban Education Reform in Abbott Districts* (No. PTM # 1501.97). Trenton.
- Singer, M. (2002). *An Outcomes Based Study of Whole School Reform in the Abbott Districts: An In-Depth Study of Academic Progress at the Lower Elementary Level Based on the Fourth Grade ESPAS*. Unpublished Doctoral Dissertation, Seton Hall University, South Orange.
- Sizer, T. R. (1984). *Horace's Compromise: The Dilemma of the American High School*. New York: Houghton Mifflin Company.
- Teaching, T. C. F. f. t. A. o. (1988). *An Imperiled Generation Saving Urban Schools*. Princeton: Princeton University Press.
- Trump, L. J. (1977). *A School For Everyone: Design for a Middle, Junior, or Senior High School That Combines the Old and the New*. Reston: The National Association of Secondary Schools Principals.
- Trump, L. J., & Georgiades, W. (1978). *How to Change Your School*. Reston: The National Association of Secondary School Principals.
- Weiler, J. (1998). Success For All: A Summary of Evaluations. *Institute for Urban and Minority Education, Teachers College*, 139(December 1998).

APPENDICES

Appendix A

List of 30 Abbott Districts

The 30 Abbott Districts are...Asbury Park City, Bridgeton City, Burlington City, Camden City, East Orange City, Elizabeth City, Garfield City, Gloucester City, Harrison Town, Hoboken City, Irvington Township, Jersey City, Keansburg Borough, Long Branch City, Millville City, New Brunswick City, Newark City, City of Orange Township, Passaic City, Paterson City, Pemberton Township, Perth Amboy City, Phillipsburg Town, Pleasantville City, Trenton City, Union City, Vineland City, and West New York Town, and the following districts not included above but designated Abbott districts pursuant to P.L. 1999, c.110, Neptune Township and Plainfield...(Urban Education, Definitions Section). ("Abbott v. Burke," 1998)

Appendix B

Urban Public Schools ESPA/NJ ASK, GEPA and HSPA Assessment Data

ESPA/NJ ASK4 District Proficiency Levels						
Year	Partially Proficient	Proficient	Partially Proficient	Proficient	Partially Proficient	Proficient
	Lang. Arts	Lang. Arts	Mathematics	Mathematics	Science	Science
1999	79.3	20.8	70.7	29.2	46.2	53.8
2000	68.8	31.2	66.5	33.5	43.3	56.6
2001	48.8	52.0	67.0	32.1	44.9	55.2
2002	35.0	65.0	61.1	38.9	NA	NA
2003	37.6	62.4	51.8	48.2	NA	NA
2004	29.1	70.9	40.7	59.3	NA	NA

(N. J. D. o. Education, 2004b)

GEPA District Proficiency Levels						
Year	Partially Proficient	Proficient	Partially Proficient	Proficient	Partially Proficient	Proficient
	Lang. Arts	Lang. Arts	Mathematics	Mathematics	Science	Science
1999	47.4	52.6	75.9	24.1	NA	NA
2000	52.5	47.5	78.3	21.7	71.2	28.9
2001	53.7	46.3	73.5	26.5	68.7	31.3
2002	53.9	46.1	69.0	31.0	66.7	33.3
2003	46.9	53.1	68.1	31.9	60.6	39.4
2004	45.5	54.5	58.4	41.6	45.5	54.5

(N. J. D. o. Education, 2004b)

Appendix B continued

Urban Public Schools ESPA/NJ ASK, GEPA and HSPA Assessment Data

HSPT/HSPA District Passing Levels			
Year	Passing Reading	Passing Mathematics	Passing Writing
1998	43.0	46.9	62.3
1999	44.6	51.7	52.5
2000	46.7	49.5	68.5
2001	NA	NA	NA
2002	NA	32.2	60.8
2003	NA	29.0	55.6

(N. J. D. o. Education, 2004b)

Appendix C

New Jersey State/NCLB District Level Benchmarks/AYP

ESPA/NJ ASK4

Subject	Year and Actual Percentage Passing and Benchmark					
	2001- 2002 Passing	2002- 2003 Bench- mark	2002- 2003 Passing	2003- 2004 Bench- mark	2004- 2005 Bench- mark	2005- 2006 Bench- mark
Lang. Arts	51.9%	58.0%	62.4%	68.0%	75.0%	75.0%
Math	32.1%	59.0%	48.2%	53.0%	62.0%	62.0%
Subject	Year and Actual Percentage Passing and Benchmark					
	2006- 2007 Bench- mark	2007- 2008 Bench- mark	2008- 2009 Bench- mark	2009- 2010 Bench- mark	2010- 2011 Bench- mark	2011- 2012 Bench- mark
Lang. Arts	75.0%	82.0%	82.0%	82.0%	91.0%	91.0%
Math	62.0%	73.0%	73.0%	73.0%	85.0%	85.0%

ESPA/NJ ASK4

Subject	Year and Actual Percentage Passing and Benchmark	
	2012- 2013	2013- 2014
	Passing	Bench- mark
Lang. Arts	91.0%	100.0%
Math	85.0%	100.0%

(T. N. P. Schools, 2003)

Appendix C

New Jersey State/NCLB District Level Benchmarks/AYP

GEPA

Subject	Year and Actual Percentage Passing and Benchmark					
	2001- 2002	2002- 2003	2002- 2003	2003- 2004	2004- 2005	2005- 2006
	Passing	Bench- mark	Passing	Bench- mark	Bench- mark	Bench- mark
Lang. Arts	46.1%	50.6%	44.3%	58%	66%	66%
Math	31.0%	36.8%	27.1%	39%	49%	49%

Subject	Year and Actual Percentage Passing and Benchmark					
	2006- 2007	2007- 2008	2008- 2009	2009- 2010	2010- 2011	2011- 2012
	Bench- mark	Bench- mark	Bench- mark	Bench- mark	Bench- mark	Bench- mark
Lang. Arts	66%	76%	76%	76%	87%	87%
Math	49%	62 %	62%	62%	79%	79%

GEPA

Subject	Year and Actual Percentage Passing and Benchmark	
	2012- 2013	2013- 2014
	Passing	Bench- mark
Lang. Arts	87%	100.0%
Math	79%	100.0%

(T. N. P. Schools, 2003)

New Jersey State/NCLB District Level Benchmarks/AYP

HSPA

Subject	Year and Actual Percentage Passing and Benchmark					
	2001- 2002	2002- 2003	2002- 2003	2003- 2004	2004- 2005	2005- 2006
	Passing	Bench- mark	Passing	Bench- mark	Bench- mark	Bench- mark
Lang. Arts	69.8%	64.1%	46.0%	73%	79%	79%
Math	32.1%	37.8%	24.2%	55%	64%	64%

Subject	Year and Actual Percentage Passing and Benchmark					
	2006- 2007	2007- 2008	2008- 2009	2009- 2010	2010- 2011	2011- 2012
	Bench- mark	Bench- mark	Bench- mark	Bench- mark	Bench- mark	Bench- mark
Lang. Arts	79%	85 %	85%	85%	92%	92%
Math	64%	74%	74%	74%	86%	86%

HSPA

Subject	Year and Actual Percentage Passing and Benchmark	
	2012- 2013	2013- 2014
	Passing	Bench- mark
Lang. Arts	92%	100.0%
Math	86%	100.0%

(T. N. P. Schools, 2003)

Appendix D

ESPA LAL Means for WSR Models 1999-2001

WSR Model	Year and Statistics								
	1999			2000			2001		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Accelerated									
Schools	27	185.00	9.60	27	181.78	12.66	27	205.21	10.92
America's									
Choice	17	179.39	14.25	16	181.52	17.07	16	201.74	9.10
CFL	20	186.75	11.67	20	183.68	14.69	20	206.35	11.53
CoNECT	23	192.89	12.96	21	187.06	13.54	24	208.77	11.31
Coalition of									
Essential									
Schools	28	187.07	12.53	27	188.14	14.44	27	211.60	12.89
Combined									
Small									
Groups	11	185.96	11.53	12	186.12	10.76	13	203.94	8.82
Comer	87	187.86	12.28	89	185.38	14.68	87	208.37	10.87
SFA	60	176.95	23.58	59	177.75	13.49	59	198.63	11.78
Ventures	11	183.44	11.77	12	185.64	14.36	12	205.38	11.63

(Singer, 2002, p. 107)

Appendix E

Chart of Literature Review

 Baltimore City Public Schools Study

Methodology	Used SFA students and matched controls, MANCOVA and effect Sizes.
Dependent Variables	Variables included reading scores obtained from the Woodcock Language Proficiency Battery and the Durrell Analysis of Reading Difficulty. Demographic variables were also used.
Findings	The implementation of SFA produced positive effects on reading scores. This study established the efficacy of SFA.
Gaps in Research	Research conducted by the developer. SFA model was not analyzed using the required district or state standardized assessments.

 Charleston County Schools Study

Methodology	Used SFA students and matched controls, F-test, ANCOVA and MANCOVA.
Dependent Variables	Variables included reading scores obtained from the Metropolitan Readiness Test, Stanford Achievement Test,

Cognitive Skills Assessment Program, Woodcock Reading Mastery Tests-Revised, Merrill Language Screening Test, Test of Language Development and the Durrell Test of Reading Difficulty.

Findings	The implementation of SFA produced mixed effects on reading scores.
Gaps in Research	Research conducted by the developer. SFA model analyzed using standardized assessments and developer recommended assessments.

New American Schools Study

Methodology	Used ANOVA to evaluate students of Roots & Wings (SFA), C0-NECT, Expeditionary Learning Outward Bound and Modern Red Schoolhouse school reform models.
Dependent Variables	Variables included reading and math scores obtained from the Texas Assessment of Academic Skills.
Findings	The implementation of SFA did not produce positive effects reading and math.
Gaps in Research	Research not conducted by the developer. The developer requested the analysis. SFA model analyzed using standardized assessments.

Singer Study

Methodology	Used t-tests, ANCOVA and MANCOVA to evaluate SFA, Accelerated Schools, America's Choice, CFL, Co-NECT, Coalition of Essential Schools, Combined Small Groups, Comer and Ventures.
Dependent Variables	Variables included Language Arts Literacy, Mathematics and Science scores obtained from the Elementary School Proficiency Assessment and other demographic variables.
Findings	The implementation of SFA produced positive effects on Language Arts Literacy, Mathematics and Science scores.
Gaps in Research	Research not conducted by the developer. SFA model analyzed using standardized assessments. Reading scores were not analyzed.
