EFFECTS OF THE TEXAS PRINCIPAL EXCELLENCE PROGRAM ON TEXAS
PRINCIPAL LEADERSHIP BEHAVIOR AND SCHOOL OUTCOMES

Todd P. Fouche, B.S., M.Ed.

Dissertation Prepared for the Degree

DOCTOR OF EDUCATION

UNIVERSITY OF NORTH TEXAS

May 2011

APPROVED:

Judith Atkison, Major Professor
James D. Powell, Minor Professor
Linda Stromberg, Committee Member and
Program Coordinator
Jimmy Byrd, Committee Member
Nancy Nelson, Chair of the Department of
Teacher Education and
Administration
Jerry R. Thomas, Dean of the College of
Education
James D. Meernik, Acting Dean of the
Toulouse Graduate School

The purpose of this study was to examine the leadership and school outcome effects of the Texas Principal Improvement program; which authorized the School Leadership Pilot Program under Texas Education Code 11.203. The specific research questions were:

What effects did participating in the Texas Principal Excellence Program in 2009-2010 have on participants and their schools?

1. What changes in participants' self-reported and peer-observed leader behaviors occurred between the initial assessment at the onset of the program and the final assessment once the program was completed?

2. What changes were experienced in school’s passing rate on mathematics and reading TAKS in schools having the same principal for the two years in 2008-2009 (pre-participation) and 2009-2010 (post participation).

The research used TxPEP participating principals who agreed to take part in the study. Principals and a selected group of people who worked with them completed a 360-degree leadership feedback instrument addressing nine leadership competencies at the beginning and end of the program. Paired samples $t$-tests were used to determine if changes from pre-participation to post-participation were statistically significant. When a statistically significant difference was found, effect size and confidence intervals were calculated to place the data in context. Multiple regression and propensity score matching were used to analyze TAKS data for the second question.
The study found that principals believed they were better able to lead after the conclusion of the TxPEP program and that their self-ratings were statistically higher on each of the nine Texas Principal competencies. The results of the 360-degree assessment showed that the peer group felt as if the principals had a statistically significant improvement on three of the nine principal competencies. Regression analysis showed there were no statistically significant changes in the school wide percent passing rates on math or reading TAKS after completion of the TxPEP program. Longitudinal research is recommended to help determine benefits of the program that might take longer to realize.
Copyright 2011

by

Todd P. Fouche
ACKNOWLEDGEMENTS

I would like to acknowledge the people who have helped to make this project a reality. This study could not have been completed without the love and support of my family. The biggest thank you goes to my wife, Melissa, for understanding and supporting me throughout the doctoral process; I could not have done this without your love and encouragement. Thank you to Tatum and Carter for not getting too upset when I had to go work on “the paper so I could be a doctor.” My hope was always that I could finish while you were young enough so that I could coach your teams and be the dad you deserve.

I would like to thank my parents, Phil Fouche and Marsha Knuth, for instilling a work ethic in me that allowed me to see this process through from beginning to end. I also could not have done this without my mother-in-law, Karen Rouse, who was always willing and available to be with the girls while I was working.

I had a wonderful committee who guided me through this journey. I owe a great deal of thanks to my major professor, Dr. Judith Adkison, who always gave me feedback quickly with a great balance of truth and encouragement. Thank you to Dr. Byrd, for your support with the methodology and steering me in the proper direction. Thank you Dr. Powell, for providing your comments and unique perspective from the business school. Finally, thank you Dr. Stromberg for your understanding of the editorial process and all of your work helping to make sure this paper was a success.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>TxPEP Program</td>
<td>3</td>
</tr>
<tr>
<td>Evaluation</td>
<td>4</td>
</tr>
<tr>
<td>ISLLC Standards</td>
<td>5</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>6</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>6</td>
</tr>
<tr>
<td>Research Questions</td>
<td>7</td>
</tr>
<tr>
<td>Limitations and Delimitations of the Study</td>
<td>7</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>8</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>9</td>
</tr>
<tr>
<td>Organization of the Dissertation</td>
<td>10</td>
</tr>
<tr>
<td>2. REVIEW OF LITERATURE</td>
<td>11</td>
</tr>
<tr>
<td>Evaluation</td>
<td>11</td>
</tr>
<tr>
<td>Principal Improvement</td>
<td>14</td>
</tr>
<tr>
<td>Evaluation of the School Leadership Pilot Program</td>
<td>18</td>
</tr>
<tr>
<td>Standards for School Principals</td>
<td>21</td>
</tr>
<tr>
<td>ISLLC Standards</td>
<td>22</td>
</tr>
<tr>
<td>TExES Standards</td>
<td>23</td>
</tr>
<tr>
<td>Research on the Standards</td>
<td>26</td>
</tr>
<tr>
<td>Assessing Principal Behavior/Knowledge</td>
<td>29</td>
</tr>
<tr>
<td>Principal Effectiveness</td>
<td>32</td>
</tr>
<tr>
<td>Summary</td>
<td>43</td>
</tr>
<tr>
<td>3. METHODOLOGY</td>
<td>44</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Comparison of Texas Competencies and ISLLC Standards</td>
<td>24</td>
</tr>
<tr>
<td>2.</td>
<td>Participant Characteristics</td>
<td>46</td>
</tr>
<tr>
<td>3.</td>
<td>Multiple Regression Analyses</td>
<td>58</td>
</tr>
<tr>
<td>4.</td>
<td>Competency 1: Vision</td>
<td>63</td>
</tr>
<tr>
<td>5.</td>
<td>Competency 2: Communication</td>
<td>64</td>
</tr>
<tr>
<td>6.</td>
<td>Competency 3: Ethics</td>
<td>65</td>
</tr>
<tr>
<td>7.</td>
<td>Competency 4: Curriculum</td>
<td>66</td>
</tr>
<tr>
<td>8.</td>
<td>Competency 5: Instruction</td>
<td>67</td>
</tr>
<tr>
<td>9.</td>
<td>Competency 6: Staff Evaluation and Development</td>
<td>68</td>
</tr>
<tr>
<td>10.</td>
<td>Competency 7: Decision Making</td>
<td>69</td>
</tr>
<tr>
<td>11.</td>
<td>Competency 8: Financial Management</td>
<td>71</td>
</tr>
<tr>
<td>12.</td>
<td>Competency 9: Management of Campus Physical Plant and School Safety</td>
<td>72</td>
</tr>
<tr>
<td>13.</td>
<td>Results of Propensity Score Matching</td>
<td>74</td>
</tr>
<tr>
<td>14.</td>
<td>Results of Multiple Regression: Outcome Variable = School Passing Percentage on 2010 Math TAKS Test</td>
<td>75</td>
</tr>
<tr>
<td>15.</td>
<td>Results of Multiple Regression: Outcome Variable = School Passing Percentage on 2010 Reading TAKS Test</td>
<td>76</td>
</tr>
<tr>
<td>16.</td>
<td>Summary of Statistically Significant Results from Question 1</td>
<td>78</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

This study is an evaluation of one year of a statewide program designed to improve principal leadership. Leadership is second only to classroom instruction among all school-related factors that contribute to student learning, with the total direct and indirect effects of leadership on student learning accounting for up to a quarter of total school effects (Leithwood, Seashore-Lewis, Anderson, & Wahlstrom, 2004). Waters, Marzano, and McNulty (2003) conducted a meta-analysis of over 70 studies with similar results, finding a substantial relationship between leadership and student achievement with an average effect size (expressed as a correlation) between the 2 of 0.25. Teachers and principals are undoubtedly the two largest contributors to student achievement. While there is a large body of research as well as practical application pertaining to teacher improvement, there is a dearth of published research and practical implications regarding the improvement of principals after they have completed their pre-service programs. The strong link between principal leadership and student learning suggests that researchers and practitioners should take notice to provide quality in-service preparation programs for existing school principals.

Educational researchers have extensively studied the practice of pre-service principal preparation programs. A search of ERIC® (http://www.eric.ed.gov) and Educational Research Complete™ (http://ebscohost.com) revealed 144 results related to principal preparation program in the abstract. The same search with the words principal improvement program submitted revealed only four results. It appears that the training of principals often ends when they graduate from their master’s program. Principals do not always emerge from graduate programs in school administration fully
prepared to lead a school (Gray, Fry, Bottoms, & O’Neill, 2007). In the corporate world and the military there are methodical initiatives to grow and develop managers, but education makes no such careful investment of resources in those who will lead this nations schools (National Policy Board for Educational Administration, 2001). In addition, there may be a time lapse between completions of administrator certification courses and becoming a principal. Many principals do not get their first principal assignment until many years after the completion of their preparation program, and almost all principals spend some time as an assistant principal before taking on the role of building leader. “In an environment of increasing accountability from the statehouse to the schoolhouse, this current ‘sink-or-swim, stumble through it’ approach to principal leadership development not only is counterproductive but helps explain why school reform efforts so often sputter and die out” (Gray et al., 2007, p. 9). Multiple studies have shown that the training principals typically receive in university programs and from their own districts do not do nearly enough to groom them for their roles as leaders of learning (Darling-Hammond, LaPointe, Meyerson, Orr, & Cohen, 2007). Quality in-service principal improvement programs must be established and become engrained in education. The proper evaluation of those programs is a key element to their longevity and effectiveness.

The current era of education has seen budgetary shortfalls that make evaluation more important now than ever before. Doing evaluation well matters in the real world because bad products and services cost lives and health, destroy the quality of life, and waste resources, such as taxes people pay to fund public education (Scriven, 1991). The legislature and central office administrators are faced with tough choices, being
forced to cancel or alter some programs to provide sufficient funds to begin or continue others (Worthen, Sanders, & Fitzpatrick, 1997). As stewards of public resources, school districts must make informed decisions about the programs they fund. This current trend of cost and program cutting lends itself well to the discipline of evaluation, specifically program evaluation as a tool that makes financial sense as well as good public rhetoric.

**TxPEP Program**

The focus of this study was a statewide program designed to improve principal leadership in Texas. House Bill 1, Article IV authorized the School Leadership Pilot Program under Texas Education Code 11.203. This program charged the Texas Education Agency with developing a school leadership program for principals in conjunction with a nonprofit corporation and improving leadership skills, student achievement, graduation rates, and teacher retention. Business schools and universities were to be consulted in the development of coursework, which focus on management and business training. A principal of a campus rated academically unacceptable was required to participate in the program and principals from other schools, which were not academically unacceptable schools, could chose to participate in the program.

The statute mandated that the Texas Education Agency develop and implement a school leadership pilot program for principals in cooperation with a nonprofit corporation that has substantial experience in developing best practices to improve leadership skills, student achievement, student graduation rates and teacher retention. A principal or a person interested in becoming a principal could apply for participation in the program. American Productivity and Quality Center (APQC) with the University of
Houston-Victoria School of Business received the grant award to establish the Texas Principal Excellence Program (TxPEP) (Texas Education Agency [TEA], 2010). An evaluation was done on the initial pilot program, which is discussed in detail in chapter 2.

Texas Principal Excellence Program (TxPEP) is an innovative self-paced professional development program for principals and aspiring principals in school districts and charter schools. The objective is to improve individual leadership skills of participants by providing professional development through convenient customized learning (TEA, 2010). One of the TxPEP area learning coordinators, Dr. Linda Monaco, revealed several changes that were made to the program after the initial pilot study (L. Monaco, personal communication, November 2, 2010). For the 2009-2010 academic year, the program was revamped and more activities were conducted virtually so that people could access the products/learning on their own schedule. A new 360-degree leadership instrument was developed which was based on the Texas state competencies for principals. Five area-learning coordinators were appointed who helped oversee the activities of participants in five different areas of the state; they also helped individualize learning for each participant.

Evaluation

Educational evaluation is the process of making decisions pertaining to the worth of educational programs (Gall, Gall, & Borg, 2007). In 1965 the United States (U.S.) government mandated that all educational programs receiving federal funding must spend a portion of that funding on evaluation; this era was the birth of educational
evaluation (Gall et al., 2007). Evaluation uses inquiry and judgment methods to
determine standards for judging quality, collecting relevant information, and applying the
standards to determine if the program that is being evaluated has merit (Worthen et al.,
1997). Scriven (1991) said the following about evaluation:

This intellectual process of evaluation is one that technology and science share
with all other disciplines, with the crafts, and with rational thought in general. It is
the process whose duty is the systematic and objective determination of merit,
worth, or value; without such a process, there is no way to distinguish the
worthwhile from the worthless. (p. 4)

ISLLC Standards

The TxPEP program was designed to address leadership competencies for
principals developed by the Texas State Board for Education Certification. The Texas
standards were developed to reflect the national standards. In the 1990s, under the
direction of the Council of Chief State School Officers (CCSSO), a group of educational
partners created the Interstate School Leaders Licensure Consortium (ISLLC) to
provide a blueprint to rebuild educational leadership in the U.S. (Murphy, 2005).
According to the CCSSO website the ISLLC standards are based on 83 empirical
studies as well as 47 sources of knowledge references (Council of Chief State School
Officers [CSSO], 1996). The desired outcome of these standards listed on the CCSSO
website is “Effective Instructional Leadership that positively impacts student
achievement” (CCSSO, 2008, p. 11). These ISLLC standards have guided the creation
of leadership policy and practice in more than 40 states (CCSSO, 1996). When
comparing the ISLLC standards with the Texas Principal Competencies it is easy to see
that Texas is one of the 40 states where the ISLLC standards served as a guide for creation.

In the fall of 2002 the Texas State Board for Educator Certification (SBEC) developed new standards for Texas Educators. The standards are based on three domains, which are further divided into nine competencies. The three domains are School and Community Leadership, Instructional Leadership, and Administrative Leadership. These three domains and nine competencies were the basis for the 360-degree observation instrument used to evaluate the performance of the principals in the TxPEP program.

Statement of the Problem

The research problem was to determine the effect of participating in the 2009-2010 TxPEP on leadership behavior of the participating principals and on student learning at their schools. Principal leadership is defined by the nine Texas Examination of Educator Standards (TExES) principal competencies. Information pertaining to principal’s specific leadership styles was obtained from a 360 degree observation instrument created by Price and Ryster (2009).

Purpose of the Study

The purpose of this study was to examine the effect of House Bill 1, Article IV, which authorized the School Leadership Pilot Program under Texas Education Code 11.203. The results of this study could provide information to entities that have a vested interest in the remediation and professional development of principals who lead academically unacceptable schools. The information gained could help guide the
Legislature while they decide the parameters of future school leadership programs. The outcomes may also help colleges and universities determine the curricular needs for their principal preparation programs.

Research Questions

The research questions guided this study:

What effects did participating in the Texas Principal Excellence Program in 2009-2010 have on participants and their schools?

1. What changes in participants' self-reported and peer-observed leader behaviors occurred between the initial assessment at the onset of the program and the final assessment once the program was completed?

2. What changes were experienced in school's passing rate on mathematics and reading Texas Assessment of Knowledge and Skills (TAKS) in schools having the same principal for the two years in 2008-2009 (pre-participation) and 2009-2010 (post participation)?

Limitations and Delimitations of the Study

Since this study focused on principal leadership and school outcomes, TAKS percent passing data were collected at the campus level; the study did not obtain individual teacher or student level data. Data collection was limited to the state of Texas. The surveys relied on principal and teacher opinions and positive or negative bias could have affected the results. The study examined data and survey data for the year prior to participation in the TxPEP program as well as immediately after the TxPEP program was over; this may not have given the participants ample time to implement the leadership skills that they obtained while participating in the TxPEP program.
Significance of the Study

This study contributed to the existing knowledge base of principal leadership and school improvement. Principals are second to teachers regarding their impact on student performance, yet there have been few studies that relate principal leadership to student achievement. By obtaining data on each of the Texas principal competencies, the information gained as a result of this study should have an impact on pre and in-service principal improvement programs. The current emphasis on educational accountability makes proficiency at the principal level more important than ever. This emphasis on accountability also makes proper program evaluation more important than ever. Evaluation can provide an objective measure to determine the merit of future and current educational programs, lack of evaluation, or worse, poor evaluation can be detrimental.

This study was different from the Learning Point study, which studied the pilot program in several key ways. The program changed significantly after the initial year, which was the year the Learning Point researchers did their evaluation; this study explored the newer version of the program. Since this study evaluated the 2009-2010 program, the data were from a developed program versus the pilot program Learning Point evaluated. The 360-degree instrument, which was used in this study, was developed and extensively studied for validity and reliability. The instrument assessed each individual on knowledge and application (scale of 1-4) on items related specifically to the nine TExES competencies, which are based on the national ISLLC standards. This study more closely examined school wide passing percentages on the TAKS test from year to year, and it was able to the track principals’ schools performance over a
longer period of time to determine if the effects of the information learned in the TxPEP program was effective. The current study systematically compared the results of the schools that the principals led, before and after those principals participated in the TxPEP program.

Definition of Terms

- **Academically unacceptable** - A School will be rated academically unacceptable if students (or groups of students) fail to meet the minimum standard on Reading/ELA, writing, social studies, math, or science TAKS tests. Student groups are compromised of African American, Hispanic, White, and Economically Disadvantaged. Other base indicators that may cause a school to be rated academically unacceptable are completion rate and dropout rate (TEA, 2010)

- **At-risk** - A set of 13 indicators that identify whether a student is more likely than a typical student to drop out of school (TEA, 2010)

- **CCSSO, Council of Chief State School Officers** - A group of educational partners who put together the Interstate School Leaders Licensure Consortium (ISLLC). Their desired outcome of ISLLC is: Effective Instructional Leadership that positively impacts student achievement (CCSSO, 1996)

- **ISLLC, Interstate School Leaders Licensure Consortium** - A set of standards developed by the CCSSO that created the vision of leadership and guided development of standards that emphasize leading for student learning, which is central to school improvement and expectations in No Child Left Behind (NCLB) (CCSSO, 1996)

- **Principal** - In this study the principal is the administrator in charge of a public school

- **SLPP, The School Leadership Pilot Program** - This program mandates that the Texas Education Agency develop and implement a school leadership pilot program for principals in cooperation with a nonprofit corporation that has substantial experience in developing best practices to improve leadership skills, student achievement, student graduation rates and teacher retention (TEA, 2010)
• Socio-economic status - For this study socio-economic status describes the economic level of the students in the school. In this study it is operationalized as the percentage of students who qualify for free or reduced lunches

• TExES -Texas Examination of Educator Standards

• TxPEP, Texas Principal Excellence Program - TxPEP is an innovative self-paced professional development program for principals and aspiring principals in school districts and charter schools. The objective is to enhance individual leadership skills of participants by providing both business management and educational administration professional development through convenient, innovative and customized learning (TEA, 2010)

• TAKS, Texas Assessment of Knowledge and Skills - TAKS is a comprehensive testing program for public school students in grades 3-11. The TAKS is designed to measure to what extent a student has learned, understood, and is able to apply the important concepts and skills expected at each tested grade level (TEA, 2010)

Organization of the Dissertation

This study is organized into 5 chapters. Chapter 1 provides an introduction, overview of key concepts, statement of the problem, purpose of the study, research questions, limitations, delimitations, definition of terms, and organization of the study. Chapter 2 is a review of the relevant literature. Chapter 3 explains the methods that were used in the research. Chapter 4 provides a presentation of the results and analysis of the data. Chapter 5 explains the results of the study and provides recommendations for future studies.
CHAPTER 2
REVIEW OF LITERATURE

The purpose of this study was to examine the effect of House Bill 1, Article IV, which authorized the School Leadership Pilot Program under Texas Education Code 11.203. The literature review begins by offering a definition for and an explanation of the discipline of evaluation. This review examined the existing information pertaining to principal improvement programs and their effectiveness. The review explored an existing evaluation of the School Leadership Pilot program and explained how this program differs. The review delved into the Interstate School Leaders Licensure consortium (ISLLC) and Texas Examination of Educator Standards (TExES), comparing the two sets of standards and exploring criticisms of the standards. Additionally, the review concentrated on the 360-degree instrument that was used to assess principal behavior and knowledge. Finally, the review investigated principal effectiveness as it relates to the TExES standards and review studies that relate effectiveness to student outcomes.

Evaluation

The methodology utilized to assess the Texas Principal Excellence Program (TxPEP) program was the methodology of evaluation. Scriven (2001) states “so there is no science without evaluation, because without evaluation one could not distinguish science from pseudo-science, let alone good science from poor science” (p. 303). This study took an evaluative perspective in determining the merit of the TxPEP program. Evaluation is not only a discipline, but a transdiscipline serving other disciplines as well;
these transdisciplines encompass a wide range from measurement to probability and
statistics to disciplines such as logic (Scriven, 2001).

Informal evaluation has gone on since the beginning of time. Scriven (2001)
states, “The earliest craft workers for which we have record, the stone shippers, left a
track record of gradually improving quality of materials and design, at single sites and
across millennia” (p. 3). Formal evaluation was evident as early as 2000 B.C. when
Chinese officials conducted civil service exams which measured proficiency of public
officials (Worthen, Sanders, & Fitzpatrick, 1997). Socrates may have been the first to
incorporate evaluation into education, using verbally mediated evaluations as part of the
learning process (Worthen, et al, 1997). The work of Ralph Tyler on curriculum
evaluation in the 1940s brought about significant changes in educational evaluation
(Gall, Gall, & Borg, 2007). Tyler introduced curriculum objectives and the success of a
program should be judged based on how well students achieved those curricular
objectives (Gall et al., 2007). In 1965 the Senate approved the Elementary and
Secondary Improvement Act, which provided grants to needy schools; each recipient
had to file an evaluation report showing what had been accomplished with federal funds
(Worthen et al., 1997). It was after this act that the field of educational evaluation began
to flourish.

In 1966 Michael Scriven published *The Methodology of Evaluation*. The main
purpose of the paper was on curricular evaluation, but almost all points in the paper
transfer to other kinds of evaluation. Scriven (1966) defines evaluation as an activity
consisting simply of gathering and combining of performance data with a weighted set
of goal scales to yield either comparative or numerical ratings. Scriven goes on to say
that evaluation is a logical activity which is essentially similar to whatever one is trying to evaluate, coffee machines or teachers, plans for a house or plans for curriculum. While the goal of evaluation remains consistent across disciplines and activities, the role of evaluation varies greatly. The role of evaluation may help to form teacher training, curriculum development, improvement of learning theory, or help to gather data on any other type of program a school may offer (Scriven, 1966).

The Joint Committee on Standards for Educational Evaluation (1994) produced a set of program evaluation standards. The standards were organized around four important attributes of evaluation: utility, feasibility, propriety, and accuracy. Utility standards guide evaluations; these standards should be informative, timely, and influential; the audience is very important and the audience’s needs should planned for well. Feasibility standards recognize that evaluations usually are conducted in a natural setting and should be practical and economical. Propriety standards are most concerned with the ethical treatment of the subjects being evaluated. Accuracy standards determine whether an evaluation has produced sound and accurate information. The joint committee recommends these standards should be applied by becoming acquainted with the program evaluation standards, clarifying the purpose and context of the program evaluation, appropriately applying the standards to determine what should be done, and finally determining what to do with the results.

Gall et al. (2007) wrote the book *Educational Research*; they found the issue of evaluation research one of enough importance that they included it as one of the chapters in the 8th edition of their book. They pointed out that evaluation differs from other types of research in three ways. First, an evaluation study is initiated by
someone’s need for a decision to be made. The purpose of the evaluation study is to collect data that will facilitate this decision. Second, an evaluation is usually done for a specific purpose while research is done so that its results may be generalized to many other situations. Third, evaluation and research differ in their judgments of value; research is more concerned with theory while evaluation explores the merit of a particular program. Gall et al. (2007) point out several steps in evaluation, specifically program evaluation; clarify the reasons for the evaluation, select a model, identify stakeholders, determine what is to be evaluated, identify questions to be answered, develop and design a timeline, collect and analyze data, and finally report the results.

Principal Improvement

The TxPEP program is far from the first or only initiative to address principal performance and improvement. In fact, recently a bill introduced in the United States (U.S.) Senate called the Teacher and Principal Improvement Act (2010) would provide $1 billion in targeted assistance to schools to develop and support effective teachers, principals, and school leaders through implementation of research based strategies. In the state of Texas the Meadows Foundation focused on principal improvement for over 25 years. On a national front the Wallace foundation recently commissioned Stanford University to execute two sets of studies: Preparing School Leaders for a Changing World (2007) and Developing Successful Principals (2005).

Developing Successful Principals is a series of in-depth case analyses of eight highly developed pre-and in-service program models in five states and tracks the graduates into schools they lead (Davis, Darling-Hammond, LaPointe, & Meyerson,
2005). The authors of this report identified four key findings from their review of literature. The first key finding contains elements essential to good leadership, which are reflected in the ISLLC standards. These standards provide common expectations and their goal is to increase student achievement by developing teachers and implementing effective organizational practices. The second finding deals with effective program design, programs which are research based and collaborative in nature produce more effective schools. The third key finding is that there are multiple ways to develop leaders, some programs focus on leadership and management skills and others emphasize academic proficiency and instructional leadership; all avenues can be effective. The fourth and final key finding is that additional research is still needed in the areas of program governance, financing, and implementation. Davis et al. (2005) point out that there is very little evidence demonstrating whether and how the kinds of learning opportunities provided by programs enable principals to become more effective building leaders. Davis et al. (2005) state, “As a result, programs are experimenting with various combinations of curriculum, methods, and program structures hoping to enhance principal practice without the solid base of empirical research to inform their design” (p. 7).

The second part of the Developing Successful Principals, Preparing School Leaders for a Changing World, focused on eight pre-and in-service principal development programs. The focus of this literature review was on the in-service programs sponsored by Hartford (CT) School District, Jefferson County (KY) Public Schools, Region 1 in New York City, and San Diego City Schools (Darling-Hammond, LaPointe, Meyerson, Orr, & Cohen, 2007).
Their study was guided by three sets of research questions:

1. Qualities of effective programs. What are the components of programs that provide effective initial preparation and ongoing professional development for principals? What qualities and design principles are displayed in these exemplary programs?

2. Program outcomes. What are the outcomes of the programs? What are principles who have experienced these trainings able to do? Do graduates of exemplary programs demonstrate instructional and organizational leadership practices that are distinctive and that are associated with more effective schools?

3. Context of High Quality Programs. What role do state, district, and institutional policies play in developing principal development programs? How do states currently manage and fund leadership development? What are the costs of exemplary preparation and professional development programs, and how are they funded? (Darling-Hammond et al., 2007, pp. 1-2)

To understand how the existing principal programs worked, Darling-Hammond et al. (2007) and her colleagues interviewed program faculty and administrators, participants and graduates, district personnel and other stakeholders; they reviewed program documents and observed meetings, courses, and workshops. They also surveyed program participants and graduates about their preparation, practices, attitudes, comparing their responses to a national random sample of principals. Finally, they observed program graduates in their jobs as principals, interviewed teachers, and examining data on school practices and student achievement trends.

The research of Darling-Hammond et al. (2007) found that it is possible to create pre-service and in-service programs that develop principals who are able to be successful in the following: cultivating a shared vision and practices, leading instructional improvement, developing organizational capacity, and managing change. Using many data sources they determined that principals who participated in the program were better prepared, had better attitudes, and engaged in more effective
practices than those principals in the comparison groups. In addition to offering
extensive high-quality learning opportunities in-service programs offered mentoring,
networking, school visits, study groups, and peer coaching. Three features
characterized successful districts in-service efforts. The first was continuous learning
that began during pre-service preparation and continued throughout their career. The
second was learning grounded in practice, where classroom practice, supervision, and
professional development were analyzed using on the job observations and a
leadership model that included readings and discussion. The final feature that
characterized successful districts was collegial learning, which is informal learning such
as networking and talking through problems with other building leaders (Darling-
Hammond et al., 2007).

The principals in these highly effective programs were also more likely to use
standards such as ISLLC to drive change; this allowed for a standards-based approach
to strengthen the approach on school improvement and instructional leadership
programs. Robust implementation of the standards through strong, tightly related
coursework and clinical experiences, reinforced by a continuum of supports upon entry
into the career, appears to be necessary to secure transformed practices (Darling-
Hammond et al., 2007).

The Meadows Principal Improvement Program has two functioning components; both are designed to improve the instructional leadership skills of principals in Texas
(Vornberg, 1989). These two components include an in-service component which is
focused on assisting practicing principals in providing the leadership necessary to
improve instruction programs in their schools and a pre-service component which is
focused on providing skilled teachers with the leadership skills necessary to make them competent instruction oriented principals (Vornberg, 1989). While the Meadows Program does include both pre-service and in-service, most of the program is focused on pre-service. The in-service component was established to provide practicing principals with an opportunity to be exposed to outstanding ideas and information, which could assist them in making productive changes in instructional programs (Vornberg, 1989). The only portion of the Meadows Program that had an evaluative component was the pre-service component; the in-service component was not evaluated.

Evaluation of the School Leadership Pilot Program

The Texas Education Agency (TEA) contracted with Learning Point Associates to provide an evaluation of the pilot year of the TxPEP program. Learning Point Associates conducted a thorough evaluation of the TxPEP pilot program. Their aim was to address the program from both a formative and summative perspective to determine program participation, implementation, and quality (Hoogstra, Hinojosa, Drill, Swanlund, Brown-Sims, Oliva, Manzeske, & Zajano, 2008). The formative evaluation included questions such as: Who were the participants; were events well attended; did events vary with regard to principal and school characteristics; was it implemented with fidelity and quality; was the content relevant to principals needs; did principals develop more knowledge and skills; and was learning from the program incorporated in the principals’ daily operations of their school?

To address the formative questions Learning Point utilized TxPEP attendance data, interviews with TEA and American Productivity and Quality Center (APQC)
program staff and cohort consultants, a survey of cohort consultants, principal interviews and focus groups, daily checklists, and principal leadership survey results. The summative portion of the evaluation addressed many questions such as what was the impact of the program on overall leadership ability, does varying amounts of program participation affect leadership, and does the program increase teacher retention? The summative portion also addressed questions about the impact of increased leadership, such as a change in student outcomes or student achievement, or the ability to lead change in any school related factors.

To address the summative questions the following data sources were used: the principal leadership survey, the leadership practices inventory (LPI), a teacher survey of principals who were either participants or in comparison groups, and administrative data on principal, school, and student characteristics, including Texas Assessment of Knowledge and Skills (TAKS) performance.

A total of 306 principals from 291 schools participated in the 2007-2008 TxPEP program; of those principals 258 were from academically unacceptable campuses while 58 were from non-academically unacceptable campuses. There was an initial kick-off summit; attendance rates at this initial summit meeting were higher than any other meeting with 84% attending. Webinar attendance rates were lower with 60-70% attending and only 28% of the participants attended all 10 required events.

Three sources of information were utilized to obtain information on the TxPEP participant’s leadership abilities: the Principal Leadership Survey, the LPI and 21st Century Principal Assessment, and a teacher survey administered to teachers of TxPEP principals and comparison schools (Hoogstra et al., 2008). The Principal Leadership
Survey was developed by Learning Point Associates to measure six leadership areas emphasized by TEA. This survey was administered to TxPEP principals as well as a comparison group of principals. The six areas the principals, as well as comparison principals, were asked to rate themselves on was effectiveness with respect to change management, building learning communities, data-driven decision making, ethical leadership, fiscal/resource management, and school/program evaluation. The LPI and 21st Century Principal Assessment are commercially available assessments. The LPI is based on effective leadership practices and the extent to which those practices are implemented. The 21st Century Principal assessment was created by the National Association of Secondary School Principals and is aligned with ISLLC leadership standards (Hoogstra et al., 2008). The teacher survey utilized was also created by Learning Point and administered to teachers who were in schools where their principal was a TxPEP participant or one of the comparison principals. This survey gathered teachers’ perceptions of school leadership, school learning environment, opportunities for teacher collaboration and decision making, teacher retention rates, promotion/graduation rates, and student engagement and performance.

In 2008 Learning Point Associates was able to answer many of their research questions at the conclusion of the study. For all leadership areas, over 80% of principals said they were incorporating what they had learned from TXPEP in their daily leadership activities. Analysis of TxPEP and comparison principals indicated that TxPEP may have had a positive impact on those principals. TxPEP principal leadership scores increased significantly while the comparison group remained the same; however this measure is a self-rating one. The LPI and 21st Century Principal assessment showed only slight
increases for some of the leadership domains measured. Teacher ratings of TxPEP principals were significantly higher for those principals who attended a high number of TxPEP events. No evidence was found on positive program impact on student performance. While campus ratings improved from 2007 to 2008, Learning Point Associates (2008) could not clearly relate those changes to the TxPEP program because campus ratings vary a great deal statewide from year to year.

While Learning Point (2008) was able to answer many questions, there were also several limitations to their study. There were low response rates to principal and teacher questionnaires. The TAKS test that was utilized to determine student achievement was administered in March, giving the principals in the program seven months or less to experience the program. While the effective schools research suggested that improved student outcomes can be attained through strategic school organization and strong principal leadership (Heck, 1992). No one principal could be expected to change learning and thus student achievement outcomes in less than seven months. Learning Point (2008) suggests that school and student outcome data might also be collected and analyzed over time to determine whether improvements occur on these indicators at schools led by principals who participate in the TxPEP program.

Standards for School Principals

In the 1990s, under the direction of the CCSSO (Council of Chief State School Officers) a group of educational partners organized the Interstate School Leaders Licensure Consortium (ISLLC) to provide a blueprint to rebuild educational leadership in the United States (Murphy, 2005). The committee’s purpose was to give a broad outline
within which leaders could be built and through which students could achieve. The goals of these standards are listed on the CCSSO website as “Effective Instructional Leadership that positively impacts student achievement” (CCSSO, 2008, p. 11). The ISLLC Standards for School Leaders were written by representatives from states and professional associations in a partnership with the National Policy Board for Educational Administration in 1994-95, supported by grants from the Pew Charitable Trusts and the Danforth Foundation. The standards were published by the CCSSO. The vision that guided development of the standards emphasizes leading for student learning, which is key to school improvement and expectations in No Child Left Behind (NCLB). The standards define what administrators should understand and be able to do to improve schools and increase student learning and were written even before NCLB’s accountability requirements for school improvement (CCSSO, 1996). The ISLLC group determined that the standards must be based on students achieving high levels and the standards should be supported by empirical findings from studies of effective schools (Murphy, 2005). Murphy (2005) went on to say that “the organizing animus for school administrators should be student learning and that the professional spotlight should shine on outcomes in this area” (p. 16).

ISLLC Standards

According to the CCSSO website, the ISLLC Standards are based on 83 empirical studies as well as 47 sources of knowledge references (CCSSO, 1996). The ISLLC Standards have helped guide leadership policy and practice in more than 40 states since they were first released in 1996 by the ISLLC. The development of the Texas
standards utilized the ISLLC standards as a guide in the creation of their TExES competencies.

**TExES Standards**

As of the fall of 2002 the State Board for Educator Certification (SBEC) developed new standards for Texas educators. Committees of Texas educators and interested citizens developed the new test based on their interpretation of what a new principal should know. This committee was armed with developing the standards, reviewing the standards, developing the framework, conducting a content validation survey, developing and reviewing new test items, conducting pilot tests of new test items, reviewing pilot test data, administering the new TExES principal test, and finally setting a new passing standard (TEA, 2010). The examination is divided into three domains; Domain I – School and Community Leadership, which is 33% of the test; Domain II – Instructional Leadership, approximately 44% of the test; and Domain III - Administrative Leadership, approximately 22% of the test. These three domains are further divided into nine competencies.

When putting certain strands side by side one can see the similarities between the two sets of standards. The following table has the TExES and ISLLC standards matched by similarities as opposed to numerical order.
Table 1

**Comparison of Texas Competencies and ISLLC Standards**

<table>
<thead>
<tr>
<th>TExES Competencies</th>
<th>ISLLC Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competency 001: The principal shapes campus culture by facilitating the development, articulation, implementation, and steward of a vision of learning that is shared and supported by the local community.</td>
<td>Standard 1: A school administrator is an educational leader who promotes the success of all students by facilitating the development, articulation, implementation, and stewardship of a vision of learning that is shared and supported by the school community.</td>
</tr>
<tr>
<td>Competency 002: The principal knows how to communicate and collaborate with all members of the school community, respond to diverse interests and needs, and mobilize resources to promote student success</td>
<td>Standard 4: A school administrator is an educational leader who promotes the success of all students by collaborating with families and community members, responding to diverse community interests and needs, and mobilizing community resources.</td>
</tr>
<tr>
<td>Competency 003: The principal knows how to act with integrity, fairness, and in an ethical and legal manner</td>
<td>Standard 5: A school administrator is an educational leader who promotes the success of all students by acting with integrity, fairness, and in an ethical manner.</td>
</tr>
<tr>
<td>Competency 005: The principal knows how to advocate, nurture, and sustain an instructional program and a campus culture that are conducive to student learning and staff professional growth.</td>
<td>Standard 2: A school administrator is an educational leader who promotes the success of all students by advocating, nurturing, and sustaining a school culture and instructional program conducive to student learning and staff professional growth.</td>
</tr>
<tr>
<td>Competency 009: The principal knows how to apply principles of leadership and management to the campus physical plant and support systems to ensure a safe and orderly learning environment.</td>
<td>Standard 3: A school administrator is an educational leader who promotes the success of all students by ensuring management of organization, operations, and resources for a safe, efficient, and effective learning environment.</td>
</tr>
</tbody>
</table>

(TEA, 2010; CCSSO, 1996)
Table 1 shows how similar the sets of standards are to one another. There are six ISLLC standards and nine TExES standards. Five of the six ISLLC standards are represented in the nine TExES standards, and, at times, they are identical. The only ISLLC standard not represented in the TExES standards is Standard 6. There are four TExES competencies that are an addition to the ISLLC standards; those are TExES competency numbers 004, 006, 007, and 008.

Like most all change initiatives, there were some criticisms of the standards movement. In “The Standards We Need”, Waters and Kingston (2005) point out that the ISLLC standards might be too broad to be practical. Waters and Grubb (2004) note “the ISLLC standards are overwhelming in scope and provide no guidance as to which of the standards has the greatest impact on student achievement. Waters and Kingston (2005) state, “The profound and rapidly increasing changes affecting schools call for standards that define a scope of essential researched-based leadership and responsibilities that reflect what school leaders need to know and be able to achieve high levels of student achievement while, at the same time, leading the redesign of the system” (p. 15). This study by Waters and Kingston identifies three standards we need. The first standard we need is the ability to distinguish essential from important leadership responsibilities; leaders should utilize research-based studies to determine what is essential. The second area that leaders should be competent in is change leadership and how to be an effective change agent. The third standard we need is one on shared leadership, which will help to sustain current and future leaders.

Waters and Kingston are not the only critics of the ISLLC standards. English (2000) dedicated an entire paper to criticizing the standards. He makes the point that
the standards are not empirically tested; instead they are based upon ‘professional consensuses. How would anyone ever know if that professional consensus were true? English goes on to dissect specific statements in the standards such as questioning the concept of continuous improvement, vision, and diversity. English ends the article comparing the ISLLC standards to religion, saying that the authors expect practitioners and the general public to accept them on faith, not empirical evidence. English (2000) states “A national system of examination and licensure which measures faith is a chilling prospect, even if it is secular in form” (p. 166).

Research on the Standards

Kaplan, Owings, and Nunnery (2005) studied principal quality in Virginia, connecting ISLLC standards with student achievement. Their study asked three questions, (1) is there a relationship between principal quality and student achievement, (2) are lower quality principals placed in lower socioeconomic schools, and (3) does principal quality have different effects at different grade levels? The authors developed a principal quality instrument based on the ISLLC standards. The study included 160 schools and the principals and they had to have a minimum of 5 years experience. Of the 160 schools, 44 were high school, 5 were primary, 61 elementary and 50 were middle schools. The test used in the study to determine student achievement was the Virginia Standards of Learning (SOL) test. In four different grade levels: third, fifth, eighth, and of course in high schools. In both third and fifth grade, schools with principals who were rated higher on the quality index tended to have higher levels of student achievement in those schools with lower quality principals: in both third and fifth
grade these findings were statistically significant. In the eighth grade and end of course groups, the results were not statistically significant, but the relationship between principal effectiveness and student outcomes was in the expected direction indicating that the study produced practically significant results.

McKerrow, Crawford, and Cornell (2006) studied the importance of the six ISLLC standards and 42 dispositions. The specific purpose was to relate the importance of the standards and dispositions to actual practice, examine the extent to which preparation programs utilized the standards, examine the differences between younger administrators trained under ISLLC and older administrators who were not, and to determine if there was a gender effect. So the information was manageable, they collapsed the 42 dispositions into the following four categories: social justice, democratic administration, courage/risk taking, and school improvement. Three hundred and forty three Illinois principals returned the survey for a 49% return rate. Descriptive and inferential statistics were compiled, specifically means, standard deviations from those means, Pearson’s correlation coefficient, and multiple $t$-tests (McKerrow et al., 2006). The data from their study indicate a negative and statistically significant relationship between years of experience and administrator’s perception about the importance of two of the ISLLC standards, collaboration with stakeholders and acting ethically. There was also a negative and statistically significant correlation between years of experience and perceptions of social justice and school improvement. These results may suggest that younger administrators who were exposed to the ISLLC standards perceive that collaboration, ethics, social justice, and school improvement are more important to the operation of a school than older administrators who were not
exposed to the standards during their principal preparation training. Gender differences were tested as well and no differences were found.

Lindle, Stalion, and Young (2004) compared the ISLLC standards to Kentucky’s Standards and Indicators for School Improvement (SISI) to identify school improvement efforts among schools both successful and low scoring on the commonwealth’s accountability system. The method used in this study is a form of content validity known as job analysis. The study involved principals self reports and observations of their daily activities. They found that ISLLC provides a useful way of describing how principals use their time, but in this study, “ISLLC showed a weakness in not providing adequate guidance for the specific leadership practices necessary for improving instruction” (Lindle et al., 2004, p.24). They also found that ISLLC did provide some direction for curriculum planning, but not very much in the way of instructional improvement or interactions with pupils with regard to their schoolwork. Lastly, they found that the ISLLC standards provided school leaders more expectations about solving problems and addressing issues pertaining to building community and collaboration (Lindle et al., 2004).

Derrington and Sharratt (2008) sought to discover the extent to which the ISLLC standards are used in the evaluation of principals in Washington State, and to identify strengths or problems in current implementation of those standards. A survey was e-mailed to 296 Washington superintendents asking them to what degree they were familiar with the ISLLC standards. The study then focused on the 16% of superintendents who reported using the standards to evaluate principals. The reason most often cited for favorable attitudes toward the standards was the specificity of
criteria and alignment with school reform. Ninety-eight principals from the district were surveyed as well; they agreed that the greatest use of the standards in evaluation is that they perform strong alignment with school reform.

Assessing Principal Behavior/Knowledge

Multirater or 360-degree feedback systems are characterized by the evaluation of an individual’s performance by multiple raters from multiple levels, meaning peers, subordinated, or superiors (Mount, Judge, Scullen, Sytsma, & Hezlett, 1998). Three hundred sixty degree feedback may be called many different names; it is also referred to as Multirater feedback, upward appraisal, full-circle feedback, and peer review (Vinson, 1996). The fundamental premise of 360-degree feedback is that the recipient of the feedback receives additional information he or she may not have received from just a supervisor. Data gathered from multiple perspectives are more comprehensive than data gathered from only one source (Dyer, 2002). Dyer (2002) went on to identify typical raters in an educational setting as superiors, peers, subordinates, parents, community members, and students; the ratings could also include family, friends, and acquaintances or anyone else who had an interest in what is being rated.

Moore (2009) makes the point that most principal evaluation systems are poorly conceived, focusing on task oriented skills as opposed to leadership. With new national and state standards, such as ISLLC and TExES, schools should have a system that provides beneficial feedback to administrators. Currently principal evaluation contains only top-down appraisals from direct supervisors who usually acquired little or no information or feedback from teachers or any other shareholders in the community. In addition, the central office administrators who evaluate principals rarely are in the
building to observe any of the competencies. Through anonymous 360-degree feedback principals can receive honest information needed to improve their leadership and, in turn, improve their schools (Moore, 2009).

Atwater, Ostroff, Yammarino, and Fleenor (1998) explored the relationship between self ratings, other ratings and outcomes. While their paper is not specifically about education, it may give insight into the relationship between leadership behavior with outcomes including student achievement. The data came from about 1,460 managers who participated in a leadership development program; the multi-rater instrument was completed by the manager, his or her peers, and subordinates. The manager’s effectiveness was also reported by each manager’s direct supervisor. There were 1,446 managers who completed self ratings, 3,939 subordinates, 3,958 peers, and 1,012 direct supervisors of the managers. The instrument used, Benchmarks, included 16 scales (106) items that measure a variety of managerial strengths and weaknesses in areas such as leading people, building and mending relationships, and acting with flexibility. Their results clearly supported the relevance of considering both self-ratings and other ratings in explaining outcomes such as supervisor ratings of managerial effectiveness, and that other ratings are most important for explaining managerial outcomes. The authors determined their study also supported the idea that self-other agreement is most relevant to outcomes that involve human perceptions and less relevant to things like sales volume or productivity goals. This finding is especially interesting for the field of education since most everything we do involves human perception.
Atkins and Wood (2002) sought to validate 360-degree feedback programs. The study had two main goals (1) to determine which source (i.e., self, superior, peer, or subordinate) from a 360-feedback program provided the most valid predictor, and (2) to better understand the relationship between self-observer discrepancies and an independent criterion. There were 63 participants in the study who were from a large Australian company in the service sector. These 63 participants were asked to rate themselves and give a questionnaire to their supervisor, a peer, and three subordinates. The respondents completed a 46 item questionnaire measuring 11 competencies. The 46 items were rated on a 5-point Likert scale, with 1 being the lowest and 5 being the highest. Each of the 63 participants also participated in an assessment center, which acted as a control group for this study. The assessment center consisted of five exercises, assessing different aspects of the participant’s role, and the raters were specialist assessors who assessed people full time for the company. Highly trained assessors in a highly structured environment provided the assessment center ratings and their ratings were shown to be a significant indicator of job performance.

Atkins and Wood (2002) found a high degree of correlation between the assessment center scores and the aggregated 360-degree scores, which provides strong validation for the use of the 360-degree feedback survey process. They found that the ratings of supervisors provided the most valid measure of staff competency and the ratings of the participants themselves provided the weakest predictor of their own competency. The more competent the participant, the more they tended to underestimate themselves and the less competent the participants were more they overestimated themselves. Even though the supervisor’s scores had the most predictive
validity, this predictive validity was enhanced when it was combined with the results of peers and subordinates. This study provides validation evidence for using 360-degree feedback instruments to determine the true behavior of leaders so that those behaviors contribution to student achievement can be measured and studied.

Principal Effectiveness

Waters and Kingston (2005) stated that “principal effectiveness is widely and easily defined: and difficult, at best, to measure. In this era of high stakes accountability, standards for school leaders must reflect what is essential to improving student achievement” (p. 15). So, their take on school effectiveness is simple; an effective school has students who achieve at a high rate. Murphy, Elliott, Goldring, and Porter (2006) offer a framework for effective leadership. Their framework includes the following tenants:

- Effective leaders facilitate the creation of a school vision that reflects high and appropriate standards of learning, a belief in the educatibility of all students, and high levels of personal and organizational performance

- Leaders in highly productive schools have a strong orientation to and affinity for the core technology of their business – learning and teaching

- Effective leaders are also knowledgeable about and deeply involved in the school’s curricular and assessment programs

- Effective school leaders are especially skillful in creating learning organizations and fostering the development of communities of learning

- High performing school leaders have a gift for acquiring, allocating and using resources to promote student success

- Effective school leaders have an emphasis on productivity, striving to maintain school accountability
Effective school leaders are social advocates, actively manipulating the environment in the service of better education for youngsters (pp. 8-30).

Porter, Goldring, Murphy, Elliott, and Cravens (2006) state “It seems that even those well schooled in the realm of educational administration struggle with quantitatively measuring leadership effectiveness; The identification and development of effective school leaders has been significantly hampered by the paucity of technically sound tools for assessing and monitoring leadership performance” (p. 3).

Porter et al. (2006) embarked on a three-year project to develop a set of instruments to study the effectiveness of educational leadership. Their assessment system, Val-Ed, is conceptualized as a multi-component assessment system for measuring critical leadership behaviors of individual educators or teams of educators, especially in urban settings for the purpose of diagnostic analysis, performance feedback, progress monitoring, and personnel decisions. (p. 3) The instrument they developed (which is based partly on ISLLC standards) to define and measure leadership behavior is two-fold; the first dimension consists of the core components of school performance, and the second dimension is the key processes of leadership. core components of school performance consisting of the following elements: high standards for school performance, rigorous curriculum, quality instruction, culture of learning and professional behavior, connections to external communities, and systemic performance accountability. The second dimension, key processes of leadership, explains leadership behaviors that can produce the core components of school performance (Porter et al., 2006). Those key processes are planning, implementing, supporting, advocating, communicating, and monitoring.
In *Measuring Leadership*, Lashway (1999) does not tell the reader how to measure a building leader, but presents, more of a process that begins with the districts leadership needs and ends when participants begin to act on the results. *Measuring Leadership* is intended primarily for those employees with district wide responsibilities, such as someone who would evaluate a principal; this is a “how to” book for school leader selection, appraisal, and development (Lashway, 1999). The book describes measurement instruments for general leadership as well as some instruments directed specifically at educational leadership. The following are some of the instruments Lashway recognized to measure educational leadership:

- **Educational Administrative Style Diagnosis Test** – is designed to stimulate thinking about the styles of educational administrators. It consists of 56 items assessing task-orientation, relationship-orientation, and effectiveness.

- **Educational Administrator Effectiveness Profile** – is designed to help school administrators understand and develop their skills in management and leadership. It consists of a 120 item 7 point Likert scale assessment; measuring eleven specific management skills: setting goals, planning, making decisions, managing business, assessing progress, delegating responsibilities, communicating, building and maintaining relationships, demonstrating professional commitment, improving instruction, and developing staff (Lashway, 1999).

- **NASSP Assessment Centers** – Are designed to help schools determine leadership talent. This assessment process uses a variety of simulated tasks such as role plays, in basket problems, and written papers. It is typically a day long assessment, followed by a feedback session.

- **The Principal Perceiver** – is designed for selection and development of school leaders. The assessment is a structured interview designed to identify twelve key “themes” in the beliefs and behaviors of school leaders: commitment, ego drive, achiever, developer, individualized perception, relator, stimulator, team, arranger, command, discipline, and concept (Lashway, 1999, pp.87-107).

- **Professional Development Inventory** – is an individual performance assessment used for diagnosing inservice needs, preparing for licensing or recertification, or for developing a professional growth plan (Lashway, 1999, pp. 87-107).
One can easily relate Lashway’s work to the ISLLC and therefore TExAS principal competencies standards. The Educational Administrator Effectiveness profile relates directly to ISLLC Standard 3 improving the management process. The Principal Perceiver and Educational Administrative Style Diagnosis Tests can be related to most all the ISLLC standards because it aims to develop leadership and stimulate the thinking of educational administrators. The Professional Development Inventory relates more directly to the TxPEP program as it is used to determine in-service needs of the principal. The test also helps to develop plans for remediation if needed.

Leithwood and Janzi (1999) studied the relative effects of principal and teacher sources of leadership on student engagement with the school. This is relevant because increased student engagement should lead to higher student achievement, which would then lead to increased school outcomes. According to Leithwood and Jantzi (1999) “Teacher leadership may be formal or informal. Formal teacher leadership is exercised by those positions such as lead teacher, master teacher, department head, union representative, member of the school’s governance council, and mentor” (p.680). Informal teacher leadership is compromised of things such as teachers who volunteer, bring ideas, and generally take ownership of the school. They used survey data from a sample of 1,762 teachers and 9,941 students in a large Canadian school district. Leithwood and Janzi (1999) propose four mediating variables or school conditions through which leadership exists: purposes and goals, school structure and social networks, people, and organizational culture. In this study they used Family Education Culture as a moderator variable rather than socioeconomic status. Leithwood and Jantzi (1999) state “Family education culture includes the assumptions, norms, values, and
beliefs held by the family about intellectual work in general, school work in particular, and the conditions that foster both” (p. 687). Leithwood and Jantzi utilized two survey instruments, one to collect data from teachers on school conditions and leadership, and the other to collect information from students pertaining to their engagement in school. The instrument on school conditions and leadership contained items measuring 5 sets of school conditions with 142 total items. These 142 items were rates on a 5 point Likert scale with responses ranging from strongly disagree to strongly agree. The study found that teacher leadership effects were insignificant. This finding caused Leithwood and Jantzi to question the whole idea of teacher leadership, making the point that teachers may not need to be leaders; administrators should be leaders and teachers should be teachers, and marrying the two concepts would do a disservice to both of them. Principal leadership effects had a statistically significant, although not a strong effect, on student engagement. Results for the mediating variable and family education culture were significant as well, and make a good case for utilizing it as opposed to student socioeconomic status in future studies.

Waters, Marzano, and McNulty (2003) examined the effects of leadership practices on student achievement with a meta-analysis that considered over 5,000 studies, of which 70 met the researchers’ strict criteria inclusion for in this study. The 70 studies they researched involved 2,894 schools with approximately 1.1 million students and 14,000 teachers. The data from their meta-analysis demonstrate that there is a substantial relationship between school leadership and student achievement, with an average effect size (expressed as a correlation) between leadership and student achievement at .25 (Waters et al., 2003, p. 3). In addition to their findings that student
achievement positively and significantly correlated to leadership, they also found that there are 21 leadership responsibilities that are responsible for student achievement. They report each of these 21 leadership responsibilities with their average effect size in a table in their report. The two responsibilities with the most significant effect size are situational awareness and intellectual stimulation. Situational awareness occurs when the leader is aware of the details of running the school and the leader uses this information to improve the school by addressing current and potential problems (Waters et al., 2003). Intellectual stimulation occurs when the leader ensures that the faculty and staff are aware of the most current and appropriate theories and practices and the leader makes the reading and discussion of these theories and practices a regular aspect of the school’s culture (Waters et al., 2003).

Hallinger, Bickman, & Davis (1996) tried to answer the question - Do principals make a difference? The study was designed to contribute to an understanding of the role of principal leadership in school effectiveness. In Tennessee 87 schools participated in the study for the duration of the project. In these 87 schools, data was collected on context factors, personal characteristics of the principals, measures of principal leadership, in-school organizational variables, and student achievement (Hallinger et al., 1996). Their first research question explored antecedent variables that might affect principal leadership; they examined the effects of parent involvement, student SES, principal gender, and prior teaching experience. Each of these variables, with the exception of prior teaching experience has a positive, significant relationship with principal leadership. The second research question examined three sets of dependent variables: instructional climate, instructional organization, and student
achievement outcomes. They found no significant direct effect of principal leadership on student achievement, meaning that principals cannot affect student achievement alone; their effect comes through indirect means, such as their effect on teachers who then have an effect on students. Therefore, Hallinger et al. (1996) did find statistically significant indirect relationships between leadership and student achievement. Their model indicates a strong relation between instructional leadership of the principal and the existence of a clear school mission, which in turn influenced student’s opportunity to learn and teachers’ expectations for student achievement. Overall their results support the idea that principals play a key role in school effectiveness.

Silins (1994) studied transformational and transactional leadership conceptualizations of leadership and their usefulness at predicting student outcomes. Attributes of transformational leadership were: visionary, goal achievement, intellectual stimulation, individual consideration and support, collaborative problem solving, and ethos. Transactional leadership was defined by bureaucratic orientation and management-by-exception. Data for the study were obtained from a survey of 291 primary teachers from 58 South Australian schools. Partial least-squares path analysis (PLSPATH) was used to develop a path model. Neither of the transactional leadership constructs contributed to an increase of student achievement. Two of the transformational leadership behaviors were positively associated with gains in student achievement, Goal Achievement and Ethos. One third of the variance found in Student Performance ($R^2 = 0.35$) is explained by the leadership behaviors of Goal Achievement and Ethos (Silins, 1994). This was not an American study, so the ISLLC standards were not studied, but ethos deals with ethics and vision, which are reflected,
in ISLLC Standards 1 and 5. Setting school goals and achieving those goals is evident throughout the ISLLC standards.

Miller and Rowan (2006) studied the effects of organic management on student achievement. They defined organic management as:

A shift away from conventional, hierarchical patterns of bureaucratic control toward what has been referred to as a network pattern of control, that is, a pattern of control in which line employees are actively involved in organizational decision making, staff cooperation and collegiality supplant the hierarchy as a means of coordinating work flows and resolving technical uncertainties, and supportive (as opposed to directive) forms of administrative leadership emerge to facilitate line employees work. (pp. 219-220)

Miller and Rowan used two large scale data sets which contain longitudinal information on student achievement: NELS: 88 and Prospects: The Congressionally-mandated Study of Educational Opportunity. The data were collected from three cohorts of elementary students, of which, two of the cohorts were utilized. Cohort 1 mathematics sample included 5,463 students in 143 schools, while the reading sample consisted of 5,561 students in 146 schools. Cohort 3 mathematics sample consisted of 5,250 students in 137 schools and the reading sample was made up of 5,314 students in 138 schools. The data were nested within schools to test the effects of school-to-school variations in the effect of organic management on student achievement.

Using data from both Prospects and NELS: 88, Miller and Rowan (2006) created duplicate measures of organic management based upon common survey items used in teacher questionnaires from each of the studies. Three indicators of organic management were (1) a measure of supportive leadership by school administrators, (2) a measure of teachers' control over key instructional decisions, and (3) a measure of staff collaboration existing in the school. Six separate three-level hierarchical linear
modeling growth models were used to estimate the effects of organic management on growth in math and reading in both cohorts as well as growth in high school reading and math achievement. There were a total of 54 tests of the effects of organic management on student achievement. Of these 54 only 5 statistically significant effects were found, and one of those was in the wrong direction. This study strongly suggests that organic forms of management are not especially powerful predictors of student achievement. The minor exception was the effect of common planning time on achievement at both the elementary and high school levels (Miller & Rowan, 2006, p. 242).

Leithwood and Mascall (2008) studied collective leadership effects on student achievement. Specifically, their study aimed to estimate, the following:

- The impact of collective leadership on key teacher variables and on student learning
- The relative influence on school decision making of each individual or group included in our measure of collective leadership (administrators, individual teachers, groups of teachers, parents, students)
- Whether differences in the patterns of collective leadership are related to differences in student achievement levels (p. 531)

They conceptualized and measured collective leadership as a form, or subset, of distributed leadership, which also overlapped Miller and Rowans definition of organic leadership, described in the previous study. Leithwood and Mascall (2008) point out the chief limitation in Miller and Rowans work is their intentional omission of indirect effects of leadership behavior on student achievement. The framework for this study presumed indirect leadership effects and conceptualized a set of teacher performance antecedents as mediators. (p. 534)

The study utilized data from 9 states, 45 districts within those 9 states and 180
schools within those 45 districts. The final data sample utilized 2,570 teacher responses from 90 schools. The survey measured collective leadership and teacher performance antecedents described in the framework: 9 items on collective leadership, 9 items on teacher capacity, 17 items on teacher motivation, and 14 items on teacher work settings and conditions (Leithwood & Mascall, 2008). Means, standard deviations, and reliability was carried out by SPSS™ (SPSS Inc., http://www.SPSS.com) and paired t-tests compared mean ratings of various sources of leadership. Hierarchical multiple regression examined the moderating effects of student socioeconomic status (SES) and LISREL was used to test a model of the relationship among collective leadership, teacher motivation, capacity and setting, and student achievement (Leithwood & Mascall, 2008).

The results of the study showed that collective leadership is significantly related to all teacher variables. The strongest relations are with collective leadership and teachers’ work setting ($r = .58$), followed by teacher motivation ($r = .55$) and all variables other than teacher capacity are significantly related to student achievement. As a whole, collective leadership effects on teachers and students explain 20% of the variation in student achievement.

Leithwood and Jantzi (2008) explored the nature, causes and consequences of school leader efficacy, including indirect influences on student learning. The authors posed four questions at the beginning of the paper, only one of those is directly related to student learning. Leithwood and Jantzi (2008) asked, “What is the contribution of leaders efficacy to variations in student learning? School leaders' efficacy are of two types – beliefs about one’s own self-efficacy for improving instruction and student
learning (LSE) and beliefs about the collective capacity of colleagues (LCE) across schools in the district to improve student learning” (p. 498). They identified four school conditions, which would significantly contribute to student learning: school culture, decision making processes, supports for instruction, and professional learning community. Leithwood and Jantzi (2008) also identified four sets of classroom conditions with significant effects on student learning: workload, areas of formal preparation, student grouping, and curriculum and instruction.

Stratified random sampling procedures were employed to choose schools (180) within districts (45) within states (9) (Leithwood & Jantzi, 2008). Data for the paper were provided by 58 items from the principal survey (measuring district variables) and 56 items from the teacher survey (measuring school and classroom variables). Because of the larger focus on leaders’ influence on student learning, they incorporated into the principal survey a 6-item scale measuring feelings of self-efficacy about instructional leadership tasks. The authors used five types of analysis to answer different questions posed in this study: Pearson product correlations, standard multiple regression, hierarchical multiple regression, t-test, and ANOVA. Leithwood and Jantzi found that LSE was not significantly related to any of the estimates of student achievement. LCE was significantly related to student achievement in two of the three years that they looked at achievement data. When combining LCE and LSE to create an aggregate efficiency measure, all three annual scores had significant relationships. In conclusion, they found weak but significant effects of leader efficacy on student learning when defining student learning as the proportion of students in schools reaching or exceeding the state’s proficient level.
Summary

This literature review began by offering a definition and explanation for evaluation. The review then explored the existing information pertaining to principal improvement programs and their effectiveness. Next, this review identified the previous study performed on the school leadership pilot program. Then the review compared and contrasted the Texas Principal Competencies and ISLLC standards and explored criticisms of the standards. Next, the review examined existing information on 360 degree instruments. Finally the review explored principal effectiveness as it related to the Texas Principal Competency standards and reviewed studies that relate the effectiveness to student outcomes.

This review addressed the issue of the lack of in-service programs for principal improvement; the programs that do exist do not have a good system for measuring principal performance. The literature shows that principals are second only to classroom teachers in regard to student achievement. The evaluation of this principal improvement program may not only provide information to improve the program addressed in this study, but future programs as well.
CHAPTER 3
METHODOLOGY

This chapter discusses the sample, participants, instrument, and data analysis that were used in this study. The purpose of this study was to examine the effect of a one year program implemented under House Bill 1, Article IV. This law authorized the School Leadership Pilot Program under Texas Education Code 11.203. Additionally, this study looked at the 2009-2010 program that was implemented after the original pilot program. The following research questions guided this study:

What effects did participating in the Texas Principal Excellence Program (TxPEP) in 2009-2010 have on participants and their schools?

1. What changes in participants’ self-reported and peer-observed leader behaviors occurred between the initial assessment at the onset of the program and the final assessment once the program was completed?

2. What changes were experienced in school’s passing rate on mathematics and reading Texas Assessment of Knowledge and Skills (TAKS) in schools having the same principal for the two years in 2008-2009 (pre-participation) and 2009-2010 (post participation)?

Sample

Table 2 provides an overview of the participants. There were approximately 350 participants in the TxPEP program at the beginning of the 2009-2010 school year. Of those 350 participants, 213 gave permission for their information to be used for research. Of the participants who gave permission 65 were required to be in the TxPEP program because their schools were given an academically unacceptable rating in the summer of 2009. The other 148 participants volunteered to take part. Of the 213 participants who agreed to have their information used by external researchers 117
reported that they were principals and 65 reported being either an assistant or vice principal; the other 31 participants had a wide variety of titles. Some of these titles were things such as counselor, specialist, coordinator, interventionist, intern, and instructional officer.

The participants were asked to provide their total number of years experience as a principal. Each response was categorized into one of five groups (first year, 1-2 years, 3-4 years, 5-9 years, or 10+ years) to help ensure a more evenly distributed sample. Participants in their first year as well as those with one or two years experience were the two groups with the largest number of participants.

The participants also reported the number of years serving at their current campus. Each response was placed into one of five categories (first year, 1-2 years, 3-4 years, 5-9 years, or 10+ years) to help ensure a more evenly distributed sample. The group of 1-2 years experience had more participants than any other group, followed by the group of first year participants.

The participants also described the grade configuration of the schools they served. Sixty-five of the participants responded that they served in a K-5 building. Fifty-three of the participants reported that they served in a 6-8 building, 48 in 9-12, 12 in K-12, and 2 serve in K-8. Thirty-three of the participants responded that they serve in an “other” campus, not represented on the questionnaire.
Table 2

*Participant Characteristics*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Years Experience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First year</td>
<td>61</td>
<td>29%</td>
</tr>
<tr>
<td>1-2 years</td>
<td>52</td>
<td>24%</td>
</tr>
<tr>
<td>3-4 years</td>
<td>35</td>
<td>16%</td>
</tr>
<tr>
<td>5-9 years</td>
<td>48</td>
<td>23%</td>
</tr>
<tr>
<td>10 + years</td>
<td>17</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Years Experience at Current Campus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First year</td>
<td>67</td>
<td>31%</td>
</tr>
<tr>
<td>1-2 years</td>
<td>77</td>
<td>36%</td>
</tr>
<tr>
<td>3-4 years</td>
<td>40</td>
<td>19%</td>
</tr>
<tr>
<td>5-9 years</td>
<td>25</td>
<td>11%</td>
</tr>
<tr>
<td>10 + years</td>
<td>4</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Type of School</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PK-5</td>
<td>65</td>
<td>28%</td>
</tr>
<tr>
<td>6-8</td>
<td>53</td>
<td>25%</td>
</tr>
<tr>
<td>9-12</td>
<td>48</td>
<td>23%</td>
</tr>
<tr>
<td>K-12</td>
<td>12</td>
<td>6%</td>
</tr>
<tr>
<td>K-8</td>
<td>2</td>
<td>1%</td>
</tr>
<tr>
<td>Other</td>
<td>33</td>
<td>15%</td>
</tr>
</tbody>
</table>
Variables Examined

Variables known to be related to school performance were used as controls for the analyses. School variables included percent economically disadvantaged students, percent African American, percent at-risk, percent Hispanic, grade level configuration, previous years TAKS math score, and previous years TAKS reading score. The school variables were obtained from the Texas Education Agency website.

*Economically disadvantaged* - A student may be identified as economically disadvantaged by the district if she or he meets any of the following criteria:

- Eligibility for free or reduced price meals under the National School Lunch and Child Nutrition Program
- From a family with an annual income at or below the official federal poverty line
- Eligible for Temporary Assistance to Needy Families (TANF) or other public assistance
- Received a Pell Grant or comparable state program of need based financial assistance (Texas Education Agency [TEA], 2010)

When assessing potential research topics, one of the first questions typically asked is whether the social scientist controlled for socio-economic status (SES) (Jeynes, 2002). Researchers generally include SES in their studies because it is correlated with many other variables such as family income, parental education, and parental occupation (Tate & Gibson, 1980). Socio-economic status (SES) may also be controlled because it can often account for the variance that occurs among the observations in a study (Jeynes, 2002).
**Student ethnicity** - The Texas Education Agency (TEA) implemented the new federal standard for the collection of ethnicity and race information beginning with data collected in the 2009-2010 school year. The United States Department of Education (USDE) requires that ethnicity and race be collected separately using a specific two-part question, presented in a specific order. Both parts of the question must be answered (TEA, 2010). Part one of the questionnaire asks the students' ethnicity and only identifies if the student is Hispanic or not. Part two of the questionnaire asks for the person's race, the choices are one or more of the following regardless of ethnicity: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian/Other Pacific Islander, or White.

**At-risk** - Is an indication of whether a student is currently identified as at-risk of dropping out of school. The 13 indicators are: (1) Student is pk-3rd grade and performed unsatisfactorily on a readiness assessment; (2) Student is in grade 7-12 and failed to maintain a 70 or above in two or more core classes during a current semester or previous semester; (3) Student was not advanced from one grade to the next, other than kindergarten at the parent’s request; (4) Student did not perform satisfactorily on an exam administered under TEC Subchapter B, Chapter 39; (5) Student is a parent or pregnant; (6-7) Student has been placed in an alterative education program or expelled during the previous or current school year; (8) Student is on parole; (9) Student was previously reported to have dropped out of school; (10) Student has limited English proficiency; (11) Student is in custody of the Department of Protective and Regulatory Services; (12) Student is homeless; and (13) Student currently resides or resided in the previous year in a detention facility, substance treatment facility, emergency shelter,
psychiatric hospital, halfway house, or foster group home (TEA, 2010).

*Previous years TAKS score* - For this study, this variable was simply the percentage of students on the campus who passed the math and reading TAKS tests.

*Grade configuration of the school* - This variable distinguishes if the school was an elementary, middle school, high school, or another type of school.

The following is a brief description of the TAKS test, the statewide assessment utilized in Texas:

In 1999 the 76th Session of the Texas Legislature enacted Senate Bill 103, mandating implementation of a new statewide testing program. The new testing requirements, subsequently named the Texas Assessment of Knowledge and Skills (TAKS), were implemented in spring 2003. The TAKS test is designed to measure the extent to which a student has learned and is able to apply the defined knowledge and skills at each tested grade level. Every TAKS test is directly aligned to the Texas Essential Knowledge and Skills (TEKS) (TEA, 2010).

This study used the math and reading TAKS tests for elementary, middle and high schools. Based on the aforementioned law, all eligible students in Texas must take the reading TAKS test in Grades 3-9 and the math TAKS test Grades 3-10. There is also an exit level math test as well as an exit level English language arts test administered in Grade 11. The frequency with which the math and reading TAKS tests are given allow for inferences to be made regarding the data the tests provide. Percentage of the students passing the TAKS, on each participant’s campus was used as the measurement.

**TAKS Validity and Reliability**

The Texas assessment program works to create assessments that are fair, accurate, valid, and reliable (Texas Education Agency [TEA], 2009). In addition to
ensuring reliability and validity the state utilizes processes called scaling and equating to transform test scores so they are easier to interpret and able to be compared across administrations. Scale scores allow for direct comparisons of student performance across separate test forms and different test administrations. A scale score converts raw scores into a common “scale” (TEA, 2009). Using sophisticated statistical methods, TEA “equates” the results of different tests, enabling the comparison of scale scores across test forms and testing administrations (TEA, 2009). Utilizing scaling and equating, the State Education Agency staff expresses confidence that consistency is established from year to year and that passing and commended scores are consistent even if the test varies slightly in difficulty.

Test validity is content based and tied directly to statewide curriculum (TEA, 2009). To ensure the highest level of content validity, the process of aligning TAKS to the curriculum was carefully implemented and included review by numerous committees of Texas educators (TEA, 2009). Evidence of this TAKS test validity is based on test content, relation to statewide curriculum, educator input, test developer input, and test expert input.

In relation to statewide curriculum, item writers and reviewers for each stage of development verify the alignment of test items with objectives to ensure that items measure appropriate content (TEA, 2009). Many current and former Texas educators wrote items specifically to measure TEKS objectives; these multiple sources of expertise provide for a system of direct input from educators and offers additional evidence regarding the validity of constructed TAKS tests (TEA, 2009). The staff at TEA, as well as test developers from Educational Testing Service, Pearson Educational
Measurement, and Questar, Inc., each internally reviews items that increase the probability that item is accurately measuring the appropriate objective (TEA, 2009). Texas Education Agency, in conjunction with Pearson Educational Measurement, receives ongoing input from a panel of national testing experts regarding all plans for collecting validity evidence for the Texas assessments (TEA, 2009).

There is also criterion-related evidence of validity for TAKS which was provided in a study conducted to fulfill the Senate Bill 103 requirement that TEA implement a college readiness component as part of the TAKS (TEA, 2009). The research, called the Higher Education Readiness Component study, included two parts: a contrasting groups study and a performance data correlation study. Results of the study indicated that TAKS scale scores at the met standard performance level predicted ACT scale scores of approximately 20 for Math and 18 for English. Results of the study also indicated that the TAKS scale scores at the commended performance level predicted ACT scale scores of approximately 27 for Math and 24 for English.

Texas Assessment of Knowledge and Skills (TAKS) test reliabilities are based on internal consistency measures, in particular the Kuder-Richardson Formula 20 (KR20) for tests involving dichotomously scored (multiple-choice) items on the stratified coefficient alpha for TAKS tests involving a combination of dichotomous and polytomous (short answer and extended response) items. The TAKS internal consistency reliabilities range from 0.81 to 0.93. This reliability range is quite strong; an effect size is considered small at .20, medium at .50, and strong at .80 or greater (Sprinthall, 2000).
360-Degree Leadership Assessment

The fundamental premise of 360-degree feedback is that data gathered from multiple perspectives are more inclusive than data gathered from only one source (Dyer, 2002). According to Vinson (1996), 360-degree feedback is also referred to as multirater feedback, upward appraisal, full circle feedback, and peer review. Existing principal evaluation typically contains only top-down appraisals from direct supervisors who usually acquire little information or feedback from shareholders in the community (Moore, 2009). When evaluating the role of the principal 360-degree feedback may be especially important since a principal's superior seldom sees the principal carrying out daily duties. Atwater, Ostroff, Yammarino, and Fleenor (1998) wrote a paper to determine that relationship between self-ratings, other ratings, and outcomes. Their results clearly supported that other ratings are most important for explaining managerial outcomes. Adkins and Wood (2002) also studied 360-degree feedback programs. They found that a supervisors rating of an individual had the most predictive validity, but when combined with results of peers and subordinates, those results were enhanced.

The 360-degree leadership model developed by Price and Ryster (2009) was chosen as the assessment for this study. This tool was written specifically to measure performance indicators from the nine Texas Principal Competencies. Price and Ryster (2009) worked with two groups of participants, those who were required to be in the TxPEP and a second group of principals who were not required to be in the TxPEP program, but were chosen randomly from a list of all principals in Texas.

Price and Ryster (2009) developed a leadership assessment tool for the purpose of gathering information related to the nine principal competencies and the 360-degree
model was chosen. The purpose of the 360-degree assessments is to gather and provide feedback to a person from a wide variety of sources, which include an individual's supervisor, peers, and subordinates, as well as self. The questionnaire was written with the intent to measure the performance indicators for the nine Texas Principal Competencies. The assessment consisted of 79 items rated on a 5-point Likert scale.

Price and Ryster (2009) utilized a web based survey and worked with two samples of participants. The first was a group of principals who were required to participate in the Texas Principal Excellence Program because these principals have a campus rating of academically unacceptable, and a second group of principals who were not required to participate in the program were chosen randomly from a list of all principals in Texas. The study consisted of two rounds of self ratings; the first round consisted of 160 participants, 108 in the required sample and 52 in the non-required sample; while round two yielded 112 usable ratings.

Price and Ryster (2009) have created an instrument that appears to be reliable. The study of reliability includes information on sources of measurement error, including internal consistency, stability, interrater, and scorer. An item analysis was conducted to determine if the final items on the assessment exhibited good internal consistency. Item discrimination coefficients exhibited moderate to high values (.40 to .75), with the exception of one item. Therefore all items were retained.

Internal consistency was one of the reliability metrics studied. Internal consistency is the extent to which tests or procedures assess the same characteristic, skill or quality. It is a measure of the precision between the observers or of the
measuring instruments used in a study. When measuring internal consistency to make low inference decisions, coefficients of .70 or higher are appropriate. All coefficients on this instrument ranged from .79 to .94. Therefore the coefficients for the TxPEP 360-degree assessment demonstrate that the instrument provides good internal consistency reliability.

The 360-degree assessment also was also evaluated in terms of the standard error of measurement (SEM). The SEM is a statistic used to estimate the probable range within which an individual’s true score on a test falls (Gall, Gall, & Borg, 2007). The smaller the SEM the more confidence one can have in assessments results; in the case of this assessment, all SEMs are less than 2, which is an indication that the scores in the assessment contain little error.

The test-retest method was used to estimate the stability of the TxPEP 360-degree assessment. Stability, in this instance, refers to how the subjects answered questions on the initial and second test. The test-retest correlations ranged from .48 to .68; these correlations indicate a moderate relationship and demonstrate that the scores on the assessment have sufficient stability.

When determining if an assessment is appropriate researchers look at reliability as well as validity. The aforementioned information proves the reliability of this instrument. Gall et al. (2007) state “Validity refers to the appropriateness, meaningfulness, and usefulness of specific inferences made from test scores” (p. 657). Content validity was assured in two ways. First the items were written to reflect the indicators on the nine standards for principal certification. Second, members of an expert panel reviewed the items and gave suggestions for improvement. To provide
evidence for discriminate validity, two tests were conducted: Z score differences and logistic regression.

The Z score differences investigated the score differences between the two samples on each competency. The means of the two groups range from .332 to .724; this represents approximately a one-third to three-fourths standard deviation difference between the two groups, with the non-required sample’s self rating higher for every competency. The second method used to provide evidence for discriminate validity was logistic regression, to determine if each competency could predict group membership. The Wald statistic in-group regression was used to make this determination and each competency was entered into a separate regression model. In all cases, competencies are reliable predictors of group membership with higher raw scores on a competency more predictive of being in the non-retained group.

Selection of Control Group

Propensity score matching was used specifically to answer Question 2; it enabled comparison of TAKS scores between the experimental group and a control group to help determine if the intervention of the TxPEP program made a difference in school performance. Propensity score matching created a control group of principals with similar characteristics to those principals who participated in the TxPEP program. Propensity score matching was used in the selection of a comparison group to provide a reduction in potentially biased estimation of treatment effects. In the current study, the treatment effect is the impact of the TxPEP program. Wilde and Hollister (2007) state “A propensity score is a conditional probability of participation in a treatment; it provides a
measure of how likely someone was to have participated, had they been eligible for the program based on the characteristics of those who were selected to participate” (pp. 459-460). In this way participants and non-participants with similar characteristics, as summarized by the propensity score, can be matched to create a comparison group (Wilde & Hollister, 2007). There are drawbacks to utilizing this method of analysis. When utilizing propensity score matching it is critical to properly identify the variables that determine participation, if not properly identified there would be a biased impact estimate (Heinrich, Maffioli, & Vazquez, 2010). Many researchers argue that the control group should be located as close, geographically, to the experimental group as possible, which can be very difficult in the field of education (Wilde & Hollister, 2007).

Data Analysis

Initially, descriptive statistics including sample characteristics, means, and standard deviations were calculated, subsequently univariate, multivariate, and bivariate correlations were conducted.

Statistical Tests Applied

Research Question 1: What changes in participants’ self-reported and peer-observed leader behaviors occurred between the initial assessment at the onset of the program and the final assessment once the program was completed? This question was answered by performing paired sample t-tests. Paired sample t-tests allowed for comparison of each principal to themselves; before and after they had been through the TxPEP program. Two paired sample t-tests were administered for each of the nine
principal competencies. For example, a paired sample $t$-test was performed on competency one that compared the mean average of the principals self-reported score before the TxPEP program to the mean average of the principals self-reported score after the intervention. A second paired sample $t$-test was then performed to compare the mean average of the peer observed score before participation to the mean average of the peer observed score after the intervention.

Research Question 2: What changes were experienced in school’s passing rate on mathematics and reading TAKS in schools having the same principal for the two years in 2008-2009 (pre-participation) and 2009-2010 (post participation). Two different procedures were utilized to answer research Question 2, propensity score matching and multiple regression. Two multiple regression analyses were conducted; change in the school wide percent passing on TAKS math scores was the criterion variable ($Y$) in the first equation and change in school wide percent passing on the reading TAKS scores was the criterion variable ($Y$) in the second equation. The predictor variables in both equations were percent economically disadvantaged students ($X$), percent African American students ($X$), percent Hispanic students ($X$), School level – elementary, middle, or high school ($X$), percent at-risk ($X$), previous years percent passing on math TAKS ($X$), and previous years percent passing on reading TAKS ($X$) (see Table 3).
Table 3

Multiple Regression Analyses

<table>
<thead>
<tr>
<th>Variables</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Y = Change in Math TAKS Scores</td>
<td></td>
</tr>
<tr>
<td>Y = Change in Reading TAKS scores</td>
<td></td>
</tr>
<tr>
<td>X1 = Percent at-risk</td>
<td></td>
</tr>
<tr>
<td>X2 = Percent economically disadvantaged</td>
<td></td>
</tr>
<tr>
<td>X3 = Percent African American</td>
<td></td>
</tr>
<tr>
<td>X4 = Percent Hispanic</td>
<td></td>
</tr>
<tr>
<td>X5 = Grade configuration</td>
<td></td>
</tr>
<tr>
<td>X6 = Previous years percent passing on math</td>
<td></td>
</tr>
<tr>
<td>X7 = Previous years percent passing on reading</td>
<td></td>
</tr>
</tbody>
</table>

Propensity score matching created a control group of schools with similar demographics to the experimental group being studied. The control group was matched to the experimental group based on 2009 school demographics of percent economically disadvantaged students, percent at-risk students, percent African American students, percent Hispanic students, region where the school is located, grade configuration of the school, previous years campus percent passing on math TAKS, and previous years campus percent passing on reading TAKS. An independent samples t-test was then conducted comparing the control group and experimental groups 2010 campus reading and math TAKS scores. Since the 2009 experimental and control groups were statistically the same for all variables, any difference in the 2010 t-test results would show difference that might be attributed to the TxPEP program.
For the purpose of this study statistical power and confidence intervals were conducted to place the data in context.

The reporting of confidence intervals can be an extremely effective way of reporting results. Because confidence intervals provide information on location and precision and can often be directly used to infer significance levels, they are, in general, the best reporting strategy. The use of confidence intervals is therefore strongly recommended (American Psychological Association [APA], 2001, p. 22).

Gall et al. (2007) state “Statistical power refers to the probability that a particular test of statistical significance will lead to the rejection of a false null hypothesis” (p. 143). According to Hinkle, Wiersma, & Jurs (2003) “The confidence interval is a range of values we are confident contains the population parameter” (p. 203).

Summary

Chapter 3 began with a review of the research questions that guided this study as well as a description of the participants. The variables were presented along with a synopsis of the TAKS test, the outcome variable. The 360-degree instrument utilized in the study was described. The chapter concluded with a description of the procedure and data analyses.
CHAPTER 4
ANALYSIS OF DATA

This study was designed to evaluate one year of the Texas Principal Excellence Program (TxPEP), which is a statewide principal improvement program. To do that, the following research question was asked:

What effects did participating in the Texas Principal Excellence Program in 2009-2010 have on participants and their schools?

1. What changes in participants’ self-reported and peer-observed leader behaviors occurred between the initial assessment at the onset of the program and the final assessment once the program was completed?

2. What changes were experienced in school wide passing rate on mathematics and reading Texas Assessment of Knowledge and Skills [TAKS] in schools having the same principal for the two years in 2008-2009 (pre-participation) and 2009-2010 (post participation)?

This chapter presents the results of the data analysis for the research question. Descriptive and inferential statistics are presented for both research questions. Question 1 addressed two variables, pre and post test scores, so a paired samples $t$-test was appropriate. Question 2 included many variables; therefore multiple regression was conducted to determine which predictor variables may have caused a change in the criterion variable. Propensity score matching was also conducted to determine causal inference, after the matching group was selected, and an independent samples $t$-test was calculated to determine if there was a statistically significant difference between the control and the experimental group.
Research Question 1

The first research question asked what changes in participants’ self reported and peer observed leader behaviors occurred between the initial assessment, before the intervention of the TxPEP program, and the final assessment after the TxPEP program was completed. The variables for this question had just two categories, pre-test and post-test, so a paired sample t-test was the appropriate statistic to determine if the difference between the pre and post scores was greater than just chance. When a statistically significant difference was found at the .05 level, effect size and confidence intervals were then calculated. Effect size provides a measure of the magnitude of the difference, which can enable a researcher to determine practical significance in addition to statistical significance; a small effect size is around .25, a medium effect size around .50 and a large effect size is .80 (Hinkle, Wiersma, & Jurs, 2003). According to Hinkle et al. (2003) “Confidence intervals are a range of values that we are confident contains the population parameter” (p. 203). Because confidence intervals provide information of location and precision, they can directly infer significance levels and the American Psychological Association (APA) strongly recommends their use (APA, 2001).

Each of the following nine tables represents one of the Texas principal competencies. The score for each competency is the mean of scores on a number of component questions. Each table shows the number of participants, the mean score that participants rated themselves both before and after the TxPEP program, the mean score that peers rated the principals before and after the program, the standard deviation, t-values, p-values, as well as the effect size.
Vision

The principal shapes campus culture by facilitating a vision of learning, which is shared and supported by the community (Texas Education Agency [TEA], 2010). Table 4 shows the results of a paired samples $t$-test used to determine if the difference between pre-test and post-test scores for vision was significantly different. There was a statistically significant difference in the self rating of principals, $t = 4.443$, $p = .001$ (CI .301, .1155). The confidence interval indicates that one can be 95% confident that the mean increase after the TxPEP program would fall between .1155 and .301. In this case the mean increase was .20. It is important that the range between the confidence intervals does not contain zero, this would indicate that the two means are equal thus there would be no statistically significant effect of the intervention. The mean score for the pre TxPEP self rating was 4.1 ($SD = .43$) and the post mean score was 4.3 ($SD = .36$). This indicates that principals perceived that they became better visionaries for their school after participating in the TxPEP program. The effect size for school vision was calculated at .456, this is considered a medium effect size.

The mean peer rating actually dropped from 4.24 at the beginning of the program to 4.23 after the principals had participated in the program, indicating that the peers saw no statistically significant change in the visionary competency after the completion of the TxPEP program.
Table 4

*Competency 1: Vision*

<table>
<thead>
<tr>
<th>Intervention</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t value</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>102</td>
<td>4.10</td>
<td>.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>102</td>
<td>4.30</td>
<td>.36</td>
<td>4.443</td>
<td>.001*</td>
<td>.456</td>
</tr>
<tr>
<td>Peer Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>102</td>
<td>4.24</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>102</td>
<td>4.23</td>
<td>.39</td>
<td>-.5819</td>
<td>.5571</td>
<td></td>
</tr>
</tbody>
</table>

**Communication**

The principal knows how to communicate and collaborate with community members, respond to diverse interests and needs, and promote student success (TEA, 2010). Table 5 shows the results of a paired samples *t*-test used to determine if the difference between pre-test and post-test scores for communication was significantly different. There was a statistically significant difference in the self-rating of principals, *t* = 6.416, *p* = .016 (CI .4179, .2205). The mean score for the pre TxPEP self rating was 3.87 (*SD* = .48) and the post mean score was 4.19 (*SD* = .38) this indicates that principals perceive that they became better communicators after participation in the TxPEP program. The effect size for communication was calculated at .675, this is considered to be a medium to high effect size.

The mean peer rating did increase from 4.06 at the beginning of the program to 4.09 after the principals had participated in the program, indicating an increase, but the
change was not enough to produce statistically significant results in the peer’s perception of the principal’s communication.

Table 5

**Competency 2: Communication**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t value</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self Rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>102</td>
<td>3.87</td>
<td>.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>102</td>
<td>4.19</td>
<td>.38</td>
<td>6.416</td>
<td>.016*</td>
<td>.675</td>
</tr>
<tr>
<td><strong>Peer Rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>102</td>
<td>4.06</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>102</td>
<td>4.09</td>
<td>.41</td>
<td>1.54</td>
<td>.127</td>
<td></td>
</tr>
</tbody>
</table>

**Ethics**

The principal acts with integrity, fairness, and in an ethical and legal manner (TEA, 2010). Table 6 shows the results of the paired samples t-test used to determine if the difference between pre-test and post-test scores for ethics was significantly different. There was a statistically significant difference in the self-rating of principals, \( t = 2.5, p = .014 \) (CI .1761, .0199). The mean score for the pre TxPEP self rating was 4.56 (\( SD = .38 \)) and the post mean score was 4.66 (\( SD = .36 \)). This indicates that principals perceived that they improved ethically after participation in the program. The effect size for Ethics was relatively small, calculated at .26.
The mean peer rating actually decreased from 4.42 at the beginning of the program to 4.40 after the principals had participated in the program, indicating that the peers saw no statistically significant change in the ethics competency after the principals completed the TxPEP program.

Table 6

<table>
<thead>
<tr>
<th>Competency 3: Ethics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
</tr>
<tr>
<td>Self Rating Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Peer Rating Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
</tbody>
</table>

Curriculum

The principal knows how to design curriculum, facilitate its implementation, and promote the utilization of various assessments to increase student performance (TEA, 2010). Table 7 shows the results of the paired samples t-test used to determine if the difference between pre-test and post-test scores for curriculum was statistically significant. There was a statistically significant difference in the self-rating of principals, $t = 4.65$, $p = .001$ (CI .3397, .1365). The mean score for the pre TxPEP self rating was 4.09 ($SD = .46$) and the post test mean score was 4.32 ($SD = .39$) which indicates that principals perceived that they became better curricular leaders for their school after
participation in the TxPEP program. The effect size for curriculum was calculated at .4891, considered a “medium” effect size.

There was also a statistically significant difference within the peer rating of the principals, $t = 2.21, p = .029$ (CI .1338, .0073). The mean score for the pre TxPEP peer rating was 4.25 ($SD = .32$) and the post test mean score was 4.29 ($SD = .37$). This indicates that principal’s peers may believe that principals have an increased understanding of curriculum after having participated in the TxPEP program. The effect size of the peer rating was calculated at .1291, considered a “small” effect size.

Table 7

**Competency 4: Curriculum**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t value</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Rating</td>
<td>Pre</td>
<td>114</td>
<td>4.09</td>
<td>.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>102</td>
<td>4.32</td>
<td>.39</td>
<td>4.65</td>
<td>.001*</td>
</tr>
<tr>
<td>Peer Rating</td>
<td>Pre</td>
<td>114</td>
<td>4.25</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>102</td>
<td>4.29</td>
<td>.37</td>
<td>2.21</td>
<td>.029*</td>
</tr>
</tbody>
</table>

**Instruction**

The principal nurtures, advocates, and sustains an instructional program that is conducive to student achievement and staff professional growth (TEA, 2010). Table 8 shows the results of the paired samples $t$-test used to determine if the difference between pre-test and post-test scores for instruction was significantly different. There
was a statistically significant difference in the self-rating of principals, \( t = 4.905, p = .001 \) (CI .308, .1306). The mean score for the pre TxPEP self rating was 4.12 (\( SD = .39 \)) and the post mean score was 4.32 (\( SD = .37 \)) which indicates that principals perceived that they improved as instructional leaders after the principals had participated in the TxPEP program. The effect size for instruction was medium sized, calculated at .5163.

The mean of the peer rating did increase from 4.24 at the beginning of the program to 4.26 after the principals had participated in the program, but this change was not great enough to produce statistically significant results in the instruction competency after the conclusion of the TxPEP program.

Table 8

<table>
<thead>
<tr>
<th>Competency 5: Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Self Rating Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Other Rating Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
</tbody>
</table>

*Staff Evaluation and Development*

The principal knows how to implement staff evaluation and development to improve faculty performance, establish appropriate models for supervision and staff development, and apply the legalities of human resources management (TEA, 2010).
Table 9 shows the results of the paired samples t-tests used to determine if the
difference between pre-test and post-test scores for staff evaluation and development
was different. There was a statistically significant difference in the self-rating of
principals, $t = 5.476, p = .001$ (CI .3511, .1643). The mean score for the pre TxPEP self
rating was 4.07 ($SD = .41$) and the post test mean score was 4.32 ($SD = .39$). This
indicates that principals perceived that they became more proficient at evaluating and
developing the staff after participating in the TxPEP program.

There was also a statistically significant difference within the peer rating of the
principals, $t = 2.21, p = .029$ (CI .1447, .0078). The mean score for the pre TxPEP peer
rating was 4.11 ($SD = .34$) and the post test mean score was 4.16 ($SD = .41$) this
indicates that principals’ peers may believe that principals did increase their ability to
evaluate and develop the staff after having participated in the TxPEP program. The
effect size for the principal self -rating was .6012, which is a medium sized effect and
the effect size of the peer rating was .1195, which is a small effect.

Table 9

<table>
<thead>
<tr>
<th>Competency 6: Staff Evaluation and Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Self Rating</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Other Rating</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Decision Making

The principal ensures an effective learning environment by applying appropriate decision-making and problem solving skills (TEA, 2010). Table 10 shows the results of the paired samples t-tests used to determine if the difference between pre-test and post-test scores for decision-making was significantly different. There was a statistically significant difference in the self-rating of principals, $t = -4.944$, $p = .001$ (CI .3434, .1468). The mean score for the pre TxPEP self rating was 4.11 ($SD = .46$) and the post test mean score was 4.33 ($SD = .45$). This indicates that principals perceived that they improved their decision making abilities after participating in the TxPEP program. The effect size for decision-making was medium sized, calculated at .4796.

The mean of the peer rating did increase from 4.13 at the beginning of the program to 4.16 after the principals had participated in the TxPEP program, but this change was not great enough to produce statistically significant results.

Table 10

<table>
<thead>
<tr>
<th>Competency 7: Decision Making</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Self Rating</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Other Rating</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Financial Management

The principal understands and can effectively manage the campus budget, personnel, resource utilization, financial management, and technology use (TEA, 2010). Table 11 shows the results of the paired samples t-tests used to determine if the difference between pre-test and post-test scores for financial management was significantly different. There was a statistically significant difference in the self-rating of principals, \( t = 4.4759, p = .001 \) (CI .4167, .1715). The mean score for the pre TxPEP self rating was 4.01 (\( SD = .6 \)) and the post test mean score was 4.27 (\( SD = .43 \)) which indicates that principals perceived that they increased their knowledge of financial management after participating in the TxPEP program.

There was also a statistically significant difference within the peer rating of the principals, \( t = 3.13, p = .0023 \) (CI .1716, .0385). The mean score for the pre TxPEP peer rating was 4.14 (\( SD = .37 \)) and the post test mean score was 4.21 (\( SD = .38 \)). This indicates that principals’ peers believe that principals have an increased understanding of financial management after having participated in the TxPEP program. The effect size for the self rating was .43, a medium effect size. The effect size of the peer rating was .1732, representing a small effect size, but this was the largest effect size reported by the peer group.
Table 11

**Competency 8: Financial Management**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t value</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self Rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>114</td>
<td>4.01</td>
<td>.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>102</td>
<td>4.27</td>
<td>.43</td>
<td>4.759</td>
<td>.001*</td>
<td>.43</td>
</tr>
<tr>
<td><strong>Other Rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>114</td>
<td>4.14</td>
<td>.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>102</td>
<td>4.21</td>
<td>.38</td>
<td>3.13</td>
<td>.0023*</td>
<td>.1734</td>
</tr>
</tbody>
</table>

**Management of Campus Physical Plant and School Safety**

The principal effectively manages the campus physical plant and support systems to ensure a safe and orderly learning environment (TEA, 2010). Table 12 shows the results of the paired samples t-tests used to determine if the difference between pre-test and post-test scores for management of campus physical plant and school safety was significantly different. There was a statistically significant difference in the self-rating of principals, \( t = 2.63, \ p = .012 \) (CI .2087, .0266). The mean score for the pre TxPEP self rating was 4.3 (\( SD = .51 \)) and the post test mean score was 4.4 (\( SD = .42 \)). This indicates that principals perceived that they improved in managing the campus physical plant and school safety after participation in the TxPEP program. The effect size for instruction was small, calculated at .1955.

The mean of the peer ratings actually decreased from 4.39 at the beginning of the program to 4.36 after the principals had completed the program, indicating that the
peers saw no statistically significant change in the management of campus physical
plant and school safety competency after the completion of the TxPEP program.

Table 12

<table>
<thead>
<tr>
<th>Competency 9: Management of Campus Physical Plant and School Safety</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Self Rating</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Other Rating</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
</tbody>
</table>

In summary, principals saw positive change in themselves on each of the nine
principal competencies. Peers did see change in principals on some of the
competencies, but it wasn’t as widespread or as powerful as the change the principals
themselves reported. According to the principals, they changed the most significantly
with regards to Staff Evaluation and Development and Communication, each had an
effect size of over .6. Principals perceived management of campus physical plant and
school safety as their lowest area of growth, with an effect size of .1955.

Research Question 2

The second research question asked what changes were experienced in a
school’s passing rate on mathematics and reading TAKS in schools that had the same
principal for two years in 2008-2009 (pre-participation) and 2009-2010 (post-participation). Two different types of statistical analysis, multiple regression and propensity score matching, were utilized to determine if there was a statistically significant difference between the school wide passing rate on the reading and math TAKS. Propensity score matching allows a researcher to determine causal inference. Essentially, it identifies if a cause and effect relationship between the intervention of the TxPEP program and each school's TAKS scores. Multiple regression allows researchers to make predictive inferences, and in this case, it allows one to determine if certain variables can predict the change in campus TAKS reading and math scores.

The first procedure discussed will be propensity score matching. This analysis matches the scores of schools where principals participated in the TxPEP program with a control group of schools whose principal did not take part in the program. The control group of schools came from a database containing all schools in the state of Texas. There were 114 schools with principals participating. For analysis, 5 were dropped due to missing values. The Match command in the statistical software program \textit{R} was utilized to match the remaining 109 experimental schools with 109 control group schools using the following variables: percent at-risk, percent economically disadvantaged, percent African American, percent Hispanic, region where the school is located, grade configuration of the school, and previous year’s school-wide percent passing on math and reading TAKS. The control group was created by matching the variables from the experimental group schools from the school year 2009; therefore, the control group and experimental group schools had essentially the same TAKS scores and demographics during 2009. The control and experimental groups were compared using an
independent samples $t$-test where 2010 math and 2010 reading TAKS scores were the output variables; since the schools were matched based on 2009 data, any difference in 2010 scores would show a change from 2009 to 2010.

For the first output variable, school wide passing rates on 2010 TAKS math test, there was an increase in the mean passing percentage of the school of 1.4278, but these results were not statistically significant ($p$-value = 0.39032, $t$-value = 0.85904). The second output variable was school wide passing rate on 2010 reading TAKS, again there was a school wide mean increase of 0.5619, but these results did not provide a statistically significant outcome ($P$-value = 0.55914, $t$-stat = 0.58411). The effectiveness of the matching algorithm was checked using the MatchBalance command in the R Match library. After matching, no variables were statistically significantly different with the smallest $p$-value = .37123.

Table 13

Results of Propensity Score Matching

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean difference</th>
<th>$t$ value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAKS Math</td>
<td>109</td>
<td>1.4278</td>
<td>0.85904</td>
<td>0.39032</td>
</tr>
<tr>
<td>TAKS Reading</td>
<td>109</td>
<td>0.5619</td>
<td>0.58411</td>
<td>0.55914</td>
</tr>
</tbody>
</table>

The second analysis conducted was multiple regression, there were two multiple regression equations applied. In the first equation the outcome variable was the schools percentage of students passing the 2010 math TAKS. The outcome variable of the second multiple regression equation was schools percentage of students passing the
2010 reading TAKS test. The variables controlled for in the multiple regression analyses were: whether they were in the control or experimental group (were they part of the TxPEP program or not), percent at risk, percent economically disadvantaged, percent African American, percent Hispanic, grade configuration of the school, and their scores on the previous year’s math and reading TAKS tests. When controlling for the abovementioned variables, whether a principal participated in the TxPEP program or not did not have a statistically significant effect on either of the two outcome variables.

Table 14

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>B</th>
<th>Std. Error</th>
<th>Beta</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participated in the TxPEP Program</td>
<td>7039</td>
<td>-0.1219</td>
<td>0.5475</td>
<td>-0.012</td>
<td>-0.223</td>
<td>0.8238</td>
</tr>
<tr>
<td>% At-risk</td>
<td>7039</td>
<td>-0.0328</td>
<td>0.0054</td>
<td>-0.055</td>
<td>-6.108</td>
<td>0.001*</td>
</tr>
<tr>
<td>% Economically Disadvantaged</td>
<td>7039</td>
<td>-0.0237</td>
<td>0.0052</td>
<td>-0.0502</td>
<td>-4.566</td>
<td>0.001*</td>
</tr>
<tr>
<td>% African American</td>
<td>7039</td>
<td>-0.0037</td>
<td>0.005</td>
<td>-0.0055</td>
<td>-0.75</td>
<td>0.4526</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>7039</td>
<td>0.0198</td>
<td>0.004</td>
<td>0.0509</td>
<td>4.961</td>
<td>0.001*</td>
</tr>
<tr>
<td>Elementary School</td>
<td>7039</td>
<td>2.545</td>
<td>0.3707</td>
<td>0.1045</td>
<td>6.865</td>
<td>0.001*</td>
</tr>
<tr>
<td>Middle School</td>
<td>7039</td>
<td>0.748</td>
<td>0.3735</td>
<td>0.0255</td>
<td>2.002</td>
<td>0.0454*</td>
</tr>
<tr>
<td>High School</td>
<td>7039</td>
<td>1.757</td>
<td>0.377</td>
<td>0.0562</td>
<td>4.663</td>
<td>0.001*</td>
</tr>
<tr>
<td>Previous year math score</td>
<td>7039</td>
<td>0.705</td>
<td>0.0099</td>
<td>0.0072</td>
<td>71.128</td>
<td>0.001*</td>
</tr>
<tr>
<td>Previous year reading score</td>
<td>7039</td>
<td>0.0911</td>
<td>0.0175</td>
<td>0.1545</td>
<td>5.244</td>
<td>0.001*</td>
</tr>
</tbody>
</table>
The results of a multiple regression analysis (see Table 14) did not show a statistically significant relationship between principal participation in the TxPEP program and the outcome variable of 2010 campus level math TAKS score. The analysis also did not show a statistically significant relationship between percent African American students and 2010 campus level math TAKS scores, not surprisingly all other covariates did show statistically significant relationships. The $R^2$ for the analysis above is 0.7829, which indicates that over 78% of the variance in TAKS math scores can be attributed to these predictor variables. The adjusted $R^2$ is 0.7826, only slightly different than the unadjusted $R^2$.

Table 15

**Results of Multiple Regression: Outcome Variable = School Passing Percentage on 2010 Reading TAKS Test**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>$B$</th>
<th>Std. Error</th>
<th>Beta</th>
<th>$t$ value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participated in the TxPEP Program</td>
<td>7039</td>
<td>-.1238</td>
<td>.3962</td>
<td>-.0021</td>
<td>-0.313</td>
<td>.7547</td>
</tr>
<tr>
<td>% At-Risk</td>
<td>7039</td>
<td>-.0178</td>
<td>.0039</td>
<td>-.0493</td>
<td>-4.586</td>
<td>.001*</td>
</tr>
<tr>
<td>% Economically Disadvantaged</td>
<td>7039</td>
<td>-.044</td>
<td>.0038</td>
<td>-.1523</td>
<td>-11.71</td>
<td>.001*</td>
</tr>
<tr>
<td>% African American</td>
<td>7039</td>
<td>-.0007</td>
<td>.0036</td>
<td>-.0016</td>
<td>-0.184</td>
<td>.8537</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>7039</td>
<td>.0056</td>
<td>.0029</td>
<td>.0234</td>
<td>1.932</td>
<td>.0534</td>
</tr>
<tr>
<td>Elementary School</td>
<td>7039</td>
<td>-.5918</td>
<td>.2682</td>
<td>-.0397</td>
<td>-2.206</td>
<td>.0274*</td>
</tr>
<tr>
<td>Middle School</td>
<td>7039</td>
<td>-2.095</td>
<td>.2703</td>
<td>-.1167</td>
<td>-7.752</td>
<td>.001*</td>
</tr>
<tr>
<td>High School</td>
<td>7039</td>
<td>2.532</td>
<td>.2727</td>
<td>.1323</td>
<td>9.283</td>
<td>.001*</td>
</tr>
<tr>
<td>Previous year math score</td>
<td>7039</td>
<td>.1311</td>
<td>.0072</td>
<td>.0022</td>
<td>18.28</td>
<td>.001*</td>
</tr>
<tr>
<td>Previous year reading score</td>
<td>7039</td>
<td>.5457</td>
<td>.0126</td>
<td>.1511</td>
<td>43.39</td>
<td>.001*</td>
</tr>
</tbody>
</table>
The results of a multiple regression analysis (see Table 15) did not show a statistically significant relationship between principal participation in the TxPEP program and the outcome variable of 2010 campus level reading TAKS scores. There was also no statistically significant relationship between student ethnicity (African American or Hispanic students) reading TAKS scores. All other covariates did prove to have a statistically significant effect on 2010 campus level reading TAKS scores. The $R^2$-squared for the analysis above is 0.6964, which indicates that almost 70% of the variance in TAKS reading scores can be attributed to these predictor variables. The adjusted $R^2$-squared is 0.696, only slightly different than the unadjusted $R^2$-squared.

Table 16 represents a correlation matrix, which shows a comparison between each of the variables explored in the multiple regression analysis.

Summary

This chapter presented the data and analyses addressing the question of whether principal leadership improved based on (a) principal and peer assessment of leadership skills defined by the Texas Principal competencies and (b) if the potential increase of leadership may have lead to increased campus performance, defined by the math and reading TAKS test scored from 2008-2009 to 2009-2010. The first part of the question was analyzed using a paired sample $t$-test. Since there were nine different tables that helped describe Question 1, the results of the analyses from Question 1 are summarized below.
Table 16

Summary of Statistically Significant Results from Question 1

<table>
<thead>
<tr>
<th></th>
<th>Self Rating</th>
<th>Effect Size</th>
<th>Other Rating</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td>**</td>
<td>.456</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Communication</td>
<td>*</td>
<td>.675</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ethics</td>
<td>*</td>
<td>.26</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Curriculum</td>
<td>**</td>
<td>.4891</td>
<td>*</td>
<td>.1291</td>
</tr>
<tr>
<td>Instruction</td>
<td>**</td>
<td>.5162</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Staff Development</td>
<td>**</td>
<td>.6012</td>
<td>*</td>
<td>.1195</td>
</tr>
<tr>
<td>Decision Making</td>
<td>**</td>
<td>.4796</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Financial Management</td>
<td>**</td>
<td>.43</td>
<td>*</td>
<td>.1734</td>
</tr>
<tr>
<td>Mgmt of physical plant and safety</td>
<td>*</td>
<td>.1955</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Significant at the 0.01 level.
*Significant at the 0.05 level.

As Table 16 shows, the principals rated themselves favorably after they had participated in the TxPEP principal improvement program. When principals rated themselves on each of the nine Texas principal competencies, results of the paired samples t-tests showed statistical significance. The peer group did not perceive principal growth at the levels the principals perceived themselves to grow. Statistically significant growth was found in only three of the nine competencies. The three competencies that the peer group found to be statistically significant were curriculum, staff development, and financial management.
Question 2 was analyzed utilizing two different approaches. Propensity score matching was utilized to create a control group, which was then compared to the experimental group using an independent samples \( t \)-test. A multiple regression analysis was also conducted. The following variables were utilized to create the matched group for the first analysis: percent at-risk, percent economically disadvantaged, percent African American, percent Hispanic, region where the school is located, grade configuration of the school, and previous year’s school-wide percent passing on math and reading TAKS. While there was a mean increase on both the reading and math TAKS tests, there was not a statistically significant difference between the control and experimental groups.

There were two multiple regression equations conducted, one with the outcome variable of the 2010 math TAKS test and the other with the outcome variable of the 2010 reading TAKS test. The variables controlled for in each of the multiple regression analyses were whether they were in the control or experimental group (were they part of the TxPEP program or not), percent at risk, percent economically disadvantaged, percent African American, percent Hispanic, grade configuration of the school, and their scores on the previous year’s math and reading TAKS tests.

Question 2 was specifically seeking to find if there was a statistically significant difference between the predictor variable pertaining to participation in the TxPEP program. There was no statistically significant difference on either the 2010 math or 2010 reading TAKS tests. Not surprisingly there were several statistically significant predictor variables that had a strong influence on each of the outcome variables. In the presence of all predictor variables included, the following variables had a statistically
significant influence on 2010 math scores: percent at-risk students, percent economically disadvantaged students, percent Hispanic students, school level, previous year’s TAKS math scores, and previous year’s reading scores. In the presence of all predictor variables included, the following variables had a statistically significant influence on 2010 reading scores: percent at-risk students, percent economically disadvantaged students, school level, previous year’s TAKS math scores, and previous year’s reading scores.
CHAPTER 5
SUMMARY, DISCUSSION, LIMITATIONS, RECOMMENDATIONS AND
CONCLUSIONS

The subject of principal improvement will garner more attention in Texas in the years to come. In this era of financial exigency combined with pressure to increase student learning, principals will continue to be asked to achieve better results with fewer resources. Principals and the schools they lead are going to have to become more efficient while becoming more effective. The state of Texas in conjunction with local school districts should make the improvement of those who lead schools a priority so that schools can meet all future challenges in the arena of education.

This study was designed to evaluate one year of a statewide program designed to improve principal leadership. I addressed the following research questions:

What effects did participating in the Texas Principal Excellence Program (TxPEP) in 2009-2010 have on participants and their schools?

1. What changes in participants’ self-reported and peer-observed leader behaviors occurred between the initial assessment at the onset of the program and the final assessment once the program was completed?

2. What changes were experienced in school’s passing rate on mathematics and reading Texas Assessment of Knowledge and Skills (TAKS) in schools having the same principal for the two years in 2008-2009 (pre-participation) and 2009-2010 (post participation)?

This chapter begins by presenting chapter summaries from the first four chapters, and then discusses the data and results based on the findings from chapter 4.
Following these discussions, limitations of the study are presented. The chapter ends with recommendations for future research, concluding with the significance of the study.

Discussion of the Results

As a group the principals felt their leadership ability in all 9 competencies had improved. The paired sample $t$-tests showed a statistically significant improvement for every competency. The effect sizes were small to medium for every competency but management of physical plant and safety, where the effect size was not large enough to be considered small with the criteria used. In contrast, the peer raters saw a statistically significant change for only three competencies; however the effect sizes for the changes did not even meet the criteria of .2 for small.

There are several obvious questions these results pose. Why is there such a significant difference between the results of the principal’s self-assessment and the assessment of the peers? Is self-efficacy at play? If self-efficacy increased during the program, it is possible that the principal’s beliefs changed but their behavior may not have been altered enough for others to notice. Finally, why are some changes larger than others?

The first question asks why is it that the changes in principal’s self-ratings were so much higher than the ratings of the peer group. The results were consistent with the work of London (2001), when considering many 360-degree studies he found that research supports that self ratings are typically higher than the ratings from others. The principals may have rated themselves higher because they were emotionally invested in the program. Participants were required to attend summit meetings, workshops,
meetings with learning coaches, and many webinars over a seven-month period. There was a great deal of time and energy put in by the principals. They know what they learned, so they may have a better perspective of the knowledge they gained than the peer group.

The principal’s perceived improvement may have also been higher than that of the peer groups because it is difficult for the peer group to change the impression of the principal after a relatively short period of time. Smith and Mackie (2000) state “Impressions tend to resist change, partly because an initial impression can alter the interpretation of later information; as a result, impressions may even survive the discrediting of information on which they were based” (p. 91). So, even if the principals have made great strides, the field of psychology indicates that the peer group may discredit, or not be aware of those changes. Psychologists also have identified a phenomenon called confirmation bias, which contends that once a person has made an impression of another, the original impression is difficult to change (Braynov & Sandholm, n.d.).

Another explanation is that the program helped the principals increase their sense of self-efficacy, while not significantly affecting their behavior. Self-efficacy refers to a person’s belief about his/her abilities rather than the individual’s actual skill set (Nye, 2008, p.11). This belief could eventually lead to principals becoming better leaders and also affect their school’s performance. Bandura (1997) believed that self-efficacy could affect an individual’s performance by influencing their thought process. Since principals in this study perceived a statistically significant increase on all competencies while the peers only reported those results on three competencies with
much smaller effect sizes, the tenets of self-efficacy might be at work. The principals obviously believe that they have improved; it is possible that, over time this belief in their abilities could manifest itself into improved leadership that others would also acknowledge. It is possible that the TxPEP program may have increased the principal’s sense of self-efficacy.

There could be many plausible explanations for why certain competencies seemed to change much more, or less, than others. The purpose of the TxPEP program is to increase teacher retention, graduation rates, and student achievement by improving principal’s leadership skills. This purpose might explain why the competencies in Domain III had the lowest effect sizes of any of the domains. Management of campus physical plant and school safety, financial management, and school safety are each important aspects of leading a school but the competencies in this domain do not directly affect the classroom as much as the competencies in some of the other domains. Also, the competencies in this domain were not as much of a focus for the TxPEP program.

The two competencies where the principals perceived improvement in effect size were communication and staff evaluation and development. Staff evaluation and development was also one of the three competencies where the peer group also saw a significant increase as well. The fact that the principals believed they had increased within this competency and the peer group saw it as well, bodes well for the TxPEP program’s having improved the principal’s skills pertaining to developing and evaluating the staff. Another reason for this could be that this might be an area where it is easier for the peer group and principal to see change. For example, if the principal did few to
no walk-throughs before the TxPEP program, and if he or she were in the classroom often after the program, both principals and peers could easily see a change in this competency.

Communication was the area where the principal’s perceived improvement had the largest increase effect size of any competency. While the change in the peer group was not statistically significant, the direction was positive. There are two possible explanations for the principal’s large effect size on this competency. First, like staff evaluation it could be easy for the principals to feel a change in communication. If a principal had few faculty meetings, sent few e-mails, and generally communicated little before the program, then began to communicate more frequently after the program, they would be able to quickly tell that they were communicating more often. It may take the peers longer to recognize the change; this might explain why there was a positive, but not statistically significant change with regard to the peer group. The second reason could be that principals rated themselves much lower on communication than any other competency before the TxPEP program, and this low rating allowed them a great deal more room to improve over the course of the program.

The improvement in principals’ self ratings for two competencies, Ethics and Management of Campus Physical Plant and School Safety had much smaller effect sizes scored by the principals than the others; the mean score of the peers actually dropped for these competencies as well. These competencies share some interesting commonalities that might have contributed to low effect sizes and the reduction of the mean scores. These two competencies had the two highest pre-TxPEP program peer ratings, of 4.42 and 4.39 (out of 5). This indicates that the peers thought the principals
were doing a good job on these competencies before the program. These initial high scores also leave little room for growth. Finally, it could be very difficult for someone to determine if a principal has increased in the area of ethics or management of campus physical plant and school safety. First, most all schools are required to have a safety plan currently in place, and second, a principal improvement program might not alter the ethical make up of an individual.

Even though they used different competencies and measurements, the changes reported here were fairly consistent with those found in the evaluation of the 2007-2008 pilot program (Hoogstra, Hinojosa, Drill, Swanlund, Brown-Sims, Oliva, Manzeske, & Zajano, 2008). In that evaluation the principals rated themselves on six leadership skills, and the self-evaluations increased significantly on five of those skills: change management, building learning communities, data-driven decision making, resource management, and school and program evaluation. The leadership area where the principal’s self-assessment did not show statistical significance was ethical leadership. As in the current study, the teachers’ ratings of the principals were also not as positive as the principals rating of themselves. Based on the difference between pre and post-test scores, the teachers did not rate the principals as having grown on any of the competencies explored.

The second research question asked what changes were experienced in the schools passing rate on mathematics and reading TAKS in schools having the same principal for the two years in 2008-2009 (pre-participation and 2009-2010 (post-participation). A control group was chosen based on propensity score matching. The results of an independent samples t-test, which compared the experimental group to the
matched control group, found no statistical significance for the school pass rate on either the mathematics or reading TAKS tests. A second analysis was conducted, consisting of two multiple regression analyses. While significant relationships were found between many of the predictor variables and the outcome variables, principal participation in the TxPEP program did not have a significant effect on either of the two outcome variables. While the percent passing the TAKS math and reading tests did increase for schools whose principals participated in the TxPEP program, analysis showed that the changes were not statistically significant. The regressions, which controlled school characteristics, showed that the increase in test scores in the TxPEP schools were not significantly different from those in the control group schools.

While the results of question 2 are disappointing, there may be some practical significance drawn from the independent samples t-test. The results were not statistically significant, but school-wide passing rates on the math TAKS test were 1.42 percentage points higher for principal’s schools who participated in the TxPEP program and reading scores were .5619 percentage points higher. These results may become more significant over time as principals have more time to implement and improve on the skills they learned during the program. The fact that principals themselves believe that they improved as a result of the TxPEP program and that the peers felt the principals had improved on some of the competencies, especially instructional leadership, may provide a platform for future change, thus possibly leading to increased student achievement at these leader’s schools.

Hoogstra et al. (2008) also considered math and reading TAKS scores when determining increased school outcomes while evaluating the pilot program, however
their study analyzed data at a student level versus school outcomes. When compared to a group of comparison schools for reading TAKS, participation in the TxPEP program had a negative statistically significant relationship with student performance. When they explored math TAKS scores there was no significant relationship found other than third grade math, where there was a negative statistically significant effect based on participation in the TxPEP program. In the current study there were no significant negative relationships reported, this could indicate that the changes made to the TxPEP program from the pilot year to 2009-2010 year positively effected student achievement.

All of the analyses did not provide statistically significant results; however, there were important findings in the study. The literature review revealed many studies that support the idea of principal leadership leading to increased student achievement, a lack of principal improvement programs, and the importance of 360-degree assessments in education. As previously mentioned, the data revealed that the principals universally felt better about their own abilities and the peers felt as the principals had improved on some competencies.

While the concept of 360-degree assessment isn’t new, it has not been widely adopted in the field of education. It should be noted that it provides valuable insight, even more so in a profession where an appraiser rarely sees the subordinate they are appraising such as a central office administrator evaluating a principal. Principals can receive valuable, honest information needed to improve their leadership and their schools through anonymous 360-degree feedback. Atwater, Ostroff, Yammarino, and Fleenor (1998) extensively studied self-other agreement and found that productivity is highest when both the self and other ratings are high, or when self-ratings are
substantially lower than other ratings. The results of this study fall in the former. The participants, as well as the peers in this study scored their answers on a Likert scale, which is numbered 1-5 with 5 being the highest score. All 18 t-tests that were conducted had a mean score of four or higher, indicating that both self and others rated principal performance as high.

Limitations of the Study

There are inherent concerns when conducting research that involves questionnaire responses. Many of the statistically significant results found were by the self report of principals who felt they had improved after the conclusion of the TxEPEP program. Self reports may be inflated and the peer group could have an agenda, but there were few outliers in the data and all analysis assumed that those who completed the 360-degree assessment did so accurately and honestly. Another potential limitation of the study is that the initial mean scores on the instrument were all relatively high, indicating that the principals and peers felt as if the principals were doing a good job. It is easier to increase the score of an individual who had a great deal of potential growth than someone who has little room to grow.

Another concern is that the surveys and the TAKS test were both administered at the end of the program, so principals did not have time to initiate some of the concepts they learned during the program. It is also highly possible that the principals did initiate concepts they learned, but the short time frame did not allow those programs or ideals to come to fruition.
As discussed previously in this chapter, the concept of confirmation bias creates a weakness in the peer group when determining principal improvement. “Confirmation bias connotes the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis at hand” (Nickerson, 1998, p. 175). The peer group may rely on pre-existing beliefs about the principal; therefore they may not have seen the change in the principal that the TxPEP program may have created.

The population of the study could also be a concern. There were approximately 350 participants in the TxPEP program to begin the 2009-2010 school year. Only 213 of those 350 gave permission for their results to be shared for this study. Out of those 213 that gave permission, only 117 reported being the principal of the building. It is possible that those who refused to participate had some type of commonality that might have skewed the results of the study had they agreed to participate.

Recommendations

The literature on principal improvement programs is relatively sparse, and this study adds to the existing literature by offering an evaluation of a statewide program. The evaluation of this program included an analysis of improving leadership, in the form of principals surveying themselves both before and after the program as well as the peer group rating the principals before and after the TxPEP program. The evaluation also included an analysis of school level outcomes, by comparing TAKS results before and after participants’ participation in the TxPEP program. The program had participants from all regions in the state representing different demographics, so the results of this program should be generalizable to states other than Texas.
The post-tests results, as well as the TAKS tests, were completed during the same year as the TxPEP program. Some of the habits of principals could take years to change, and the concepts taught could also take years to implement. Research indicates that the principal’s effect on student achievement is typically indirect, principals set the direction for school improvement by training and supporting the staff, and making sure the school is well managed (Leithwood & Jantzi, 1999). Longitudinal research tracking the participants’ school’s progress over time is recommended. This type of long-range study would be a better indicator of leader behavior as well as school level outcomes because it would allow the participants more time to implement their learning and it would be a better representation of the quality of the principal improvement program.

An additional recommendation is to consider outcome variables other than passing rates on the TAKS test. If longitudinal data were available on principals who participated in the program, a researcher could study TAKS scores, graduation rates, attendance patterns, teacher retention, or many other variables. Different teacher variables could be looked at as well. Is teacher attendance higher as a result of increased leadership capacity? Is quality teacher retention greater as a result of learned leadership behaviors?

Considering the principals who were required versus the principals who chose to attend the program would be another interesting area to research. This could provide information on principal motivation. The required principals are in schools with low TAKS scores, the students in their school are not performing well, but they have a great deal of room for improvement. A principal improvement such as the TxPEP program
might be more impactful on this type of principal that a principal in a school already enjoying student success.

Another potential line of research could be to determine if there are specific principal competencies or combinations of competencies that lead to increased performance. Performance could be measured either by increased leadership behavior or improvement of school wide standardized test scores. This information could be very valuable to the development or enhancement of future principal improvement programs. For example, if researchers knew that the curriculum competency and the decision making competency each predicted improvement in standardized test scores better than any other competencies, they could design principal preparation programs that focused specifically on those competencies.

Conclusion

There is not a great deal of existing research pertaining to principal improvement programs. This study adds to the literature. There is little uncertainty that the principals left the program feeling as if they were better able to lead than before the program began. This increased sense of self-efficacy could be the impetus for future school level improvements and leadership behavior enhancement.

The literature leaves no doubt that principals are second only to teachers with regard to impact on student achievement. There are a myriad of resources available to an average or struggling teacher to remediate any potential deficiency. There are few or no resources available to a principal who is not effective. The results of this study will increase the exposure pertaining to the lack of principal improvement programs and add to the knowledge base on the topic.
APPENDIX

CORRELATION MATRIX FOR MULTIPLE REGRESSION TABLE
<table>
<thead>
<tr>
<th></th>
<th>10 Math</th>
<th>TxPEP</th>
<th>At-Risk</th>
<th>ED</th>
<th>AA</th>
<th>Hisp</th>
<th>09 Math</th>
<th>09 Reading</th>
<th>10 Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Math</td>
<td>1</td>
<td>-.071</td>
<td>-.483</td>
<td>-.325</td>
<td>-.207</td>
<td>-.117</td>
<td>.88</td>
<td>.657</td>
<td>.708</td>
</tr>
<tr>
<td>TxPEP</td>
<td>-.071</td>
<td>1</td>
<td>.01</td>
<td>-.001</td>
<td>.039</td>
<td>-.009</td>
<td>-.079</td>
<td>-.042</td>
<td>-.046</td>
</tr>
<tr>
<td>At-Risk</td>
<td>-.483</td>
<td>.01</td>
<td>1</td>
<td>.698</td>
<td>.094</td>
<td>.601</td>
<td>-.502</td>
<td>-.608</td>
<td>-.556</td>
</tr>
<tr>
<td>ED</td>
<td>-.325</td>
<td>-.001</td>
<td>.698</td>
<td>1</td>
<td>.191</td>
<td>.696</td>
<td>-.313</td>
<td>-.591</td>
<td>-.587</td>
</tr>
<tr>
<td>AA</td>
<td>-.208</td>
<td>.039</td>
<td>.094</td>
<td>.191</td>
<td>1</td>
<td>-.304</td>
<td>-.203</td>
<td>-.218</td>
<td>-.212</td>
</tr>
<tr>
<td>Hisp</td>
<td>-.177</td>
<td>-.009</td>
<td>.601</td>
<td>.696</td>
<td>-.304</td>
<td>1</td>
<td>-.185</td>
<td>-.409</td>
<td>-.388</td>
</tr>
<tr>
<td>09 Math</td>
<td>.88</td>
<td>-.079</td>
<td>-.502</td>
<td>-.313</td>
<td>-.203</td>
<td>-.185</td>
<td>1</td>
<td>.711</td>
<td>.61</td>
</tr>
<tr>
<td>09 Reading</td>
<td>.657</td>
<td>-.042</td>
<td>-.608</td>
<td>-.591</td>
<td>-.218</td>
<td>-.409</td>
<td>.711</td>
<td>1</td>
<td>.799</td>
</tr>
<tr>
<td>10 Reading</td>
<td>.708</td>
<td>-.046</td>
<td>-.556</td>
<td>-.587</td>
<td>-.212</td>
<td>-.388</td>
<td>.61</td>
<td>.799</td>
<td>1</td>
</tr>
</tbody>
</table>
REFERENCES


