BENCHMARK ASSESSMENTS AND THEIR RELATIONSHIP TO THE CALIFORNIA STANDARDS TESTS, TEACHING, AND LEADERSHIP IN TITLE I ELEMENTARY SCHOOLS

Sareyrinny Hang Liemthongsamout
B.A., California State University, Sacramento, 2003
M.A., California State University, Sacramento, 2006

DISSERTATION

Submitted in partial satisfaction of the requirement for the degree of

DOCTOR OF EDUCATION

in

EDUCATIONAL LEADERSHIP

at

CALIFORNIA STATE UNIVERSITY, SACRAMENTO

SPRING
2012
BENCHMARK ASSESSMENTS AND THEIR RELATIONSHIP TO THE CALIFORNIA STANDARDS TESTS, TEACHING, AND LEADERSHIP IN TITLE I ELEMENTARY SCHOOLS

A Dissertation

by

Sareyrinny Hang Liemthongsamout

Approved by Dissertation Committee:

____________________
Edmund W. Lee, Ed. D., Chair

____________________
Robert Pritchard, Ph. D.

____________________
Mao Justice Vang, Ed. D.

SPRING 2012
BENCHMARK ASSESSMENTS AND THEIR RELATIONSHIP TO THE CALIFORNIA STANDARDS TESTS, TEACHING, AND LEADERSHIP IN TITLE I ELEMENTARY SCHOOLS

Student: Sareyrinny Hang Liemthongsamout

I certify that this student has met the requirements for format contained in the University format manual, and that this dissertation is suitable for shelving in the library and credit is to be awarded for the dissertation.

_______________________________, Director
Carlos Nevarez, Ph. D.

_______________________________
Date
DEDICATION

This research is dedicated to all underprivileged children, who have inspired me in countless ways to work relentlessly to improve education and their quality of life and future.
ACKNOWLEDGMENTS

Love for education has been my lifelong journey. This dissertation epitomizes what can be accomplished with knowledge and belief in one’s great potential, accompanied by solid ambition to achieve, and blessed with incredible, supportive people who share our dreams. I would like to recognize the many important individuals who have helped me throughout this remarkable journey.

I will be eternally grateful to my immediate family. My amazingly supportive best friend and husband Saysana Liemthongsamout and his extraordinary belief in my ability to achieve any ambition along with our loving daughters Tavi Skylar and Apsara Chloe have been generous, patient, and understanding with my time, which they shared with my pursuit of a doctoral degree and career for almost four years.

My parents, Sarinn Hang and Chhoeun Kuoch Hang, have always motivated me to strive for and achieve the highest level of excellence. They are the reason why I chose to pursue a doctoral degree. They value and believe that education is the key to all opportunities in life. This degree is a small gift to them in attempts to repay them for the many gifts they have given me.

My big brother, Sareyraksas Hang, has set the bar high from the very start and forced me to compete and strive to achieve many levels of success to avoid standing in his shadow. He is my role model and to whom I credit for instilling in me a competitive, no nonsense attitude to reach for the stars and settle for nothing less.
Wonderful friends believed in and supported me at various points in this journey: Chery Akaba-McCumber, Scott Holton, and Debi Lawless. Their interest in my studies, writing, progress, words of encouragement and unconditional beliefs in my ability to succeed are invaluable to this achievement. I would like to acknowledge Ms. Chery for spending her weekends and vacation with me at work so I would not be alone while I focused on completing my dissertation. Each of these beloved friends is special to this process in their unique ways and deserves a part of this accomplishment. It was my fear of letting them down that helped motivate me to keep pushing forward to the end.

Special recognition should be given to the California State University, Sacramento’s second cohort of doctoral students. They are a respectable group of individuals with whom I have great admiration. Particularly, I would like to thank Drs. Daryl Camp and Francine Stevens for their friendship and help with the courses, qualifying exam, and dissertation along with Drs. Gabriel Simon and Katherine Punteney, as well as Lucinda Wong for always willing to help. These people are the nicest, most generous individuals I have had the honor to meet.

Moreover, I am forever indebted to my dissertation committee Drs. Edmund W. Lee, Robert Pritchard, and Mao Justice Vang in addition to Dr. JoLynn Britt for her assistance with data analyses, support, and encouragement. Dr. Edmund W. Lee was a phenomenal professional, to whom I am eternally grateful. The happy ending to this difficult journey was possible because he genuinely cared and was invested in serving my academic and professional needs. Dr. Edmund W. Lee was also a genius in his selection
of Drs. Robert Pritchard and Mao Justice Vang to serve on the dissertation committee.

All three professionals completed each other and my dissertation with their unique perspective and expertise. Drs. Edmund W. Lee, Robert Pritchard, and JoLynn Britt exemplify the quality of California State University, Sacramento and its Doctorate in Educational Leadership program.
CURRICULUM VITAE

Education
M.A. in Curriculum and Instruction, California State University, Sacramento (2006)
B.A. in Child Development, California State University, Sacramento (2003)

Professional Employment
Site Instruction Coordinator/Assistant Principal: Sacramento City Unified School District
(2010-Present)

Kindergarten/Lead Teacher, John Still Elementary School: Sacramento City Unified
School District (2007-2010)

English-Language Arts Training Specialist, John Still Elementary School: Sacramento
City Unified School District (2006-2007)

Sixth Grade Teacher, Mark Hopkins Elementary School: Sacramento City Unified

Kindergarten/First Grade Teacher, Mark Hopkins Elementary School: Sacramento City

First Grade Teacher, Mark Hopkins Elementary School: Sacramento City Unified

Third Grade Teacher, William Land Park Elementary School: Sacramento City Unified


Teacher, Child Development Center, B. Comstock Elementary School: Sacramento,
California (1999)

Teacher, Montessori Preschool: Sacramento, California (1999)

Counselor, FamiliesFirst: Davis, California (1998-1999)
Abstract

of

BENCHMARK ASSESSMENTS AND THEIR RELATIONSHIP TO THE CALIFORNIA STANDARDS TESTS, TEACHING, AND LEADERSHIP IN TITLE I ELEMENTARY SCHOOLS

by

Sareyrinny Hang Liemthongsamout

Education plays an enormous and vital role in society as well as the United States’ economy, thus warranting laws mandating schools to implement accountability measures in response to increasing concerns about American education. Districts and schools nationwide are pressured to produce academically proficient students who are well-prepared to work, live, and contribute effectively upon graduation. Consequently, many districts and schools across the nation are currently administering periodic benchmark assessments to complement the end-of-year state testing and increase student learning. Up to the present, limited study has been conducted to investigate the relationship between student scores on the final benchmark assessments and student scores on the California Standards Tests (CST) in Title I Program Improvement elementary schools. Furthermore, no study exists on the influence of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals in these low achieving elementary schools.

This study examined the relationship between benchmark assessments and the end-of-year state testing in two PI Title I schools in one school district in northern
California (NCSD). More specifically, this study investigated whether there is a significant difference between student scores on the final benchmark assessments and student scores on the annual state tests. Many conclusions can be drawn about benchmark assessments based on both the quantitative and qualitative research methods used in this study. First, according to the comparison of CAB and CST scores, evidence shows on the ELA portion that, although there were no significant difference between students’ CAB scores and CST scores, students tended to get the same result; therefore, the CAB has predictive validity with respect to the CST. Next, unlike the ELA portion, student scores on the math CAB and CST showed significant difference in six of ten subgroups (All Students, African American, Asian, Non-EL, Second Grade, and Third Grade Students). In other words, students were more likely to score higher on the CST than the CAB; possibly indicating questions on the math CAB might be more difficult than questions on the math CST or that teachers were stronger in teaching mathematics.

Teacher and principal surveys also revealed benchmark assessments play an important role in the teaching practices of teachers and instructional leadership practices of principals. Both groups surveyed agreed with the many benefits of benchmark assessments, including its impact on small group instruction, time working with specific groups of students, and differentiated instruction. Moreover, open-ended survey questions and teacher interviews yielded definite conclusions about benchmark assessments; they provide essential data that transforms instruction by (1) identifying
students’ strengths and weaknesses and help in (2) identifying teacher challenges with teaching particular standards.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>ix</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xvi</td>
</tr>
<tr>
<td>Chapter 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Impact of High Stakes Testing</td>
<td>4</td>
</tr>
<tr>
<td>Benchmark Assessments</td>
<td>6</td>
</tr>
<tr>
<td>Title I PI Schools</td>
<td>8</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>9</td>
</tr>
<tr>
<td>Theoretical Base</td>
<td>12</td>
</tr>
<tr>
<td>Operational Definitions</td>
<td>13</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>15</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>16</td>
</tr>
<tr>
<td>Conclusion</td>
<td>17</td>
</tr>
<tr>
<td>Chapter 2. REVIEW OF RELATED LITERATURE</td>
<td>19</td>
</tr>
<tr>
<td>No Child Left Behind and High Stakes Testing</td>
<td>19</td>
</tr>
<tr>
<td>Instructional Leadership</td>
<td>29</td>
</tr>
<tr>
<td>Data-Driven Decision Making</td>
<td>30</td>
</tr>
</tbody>
</table>
Assessments ........................................................................................................ 31
Conclusion ......................................................................................................... 50

3. METHODOLOGY ............................................................................................ 51
   Research Design and Approach .................................................................... 54
   Setting and Sample ....................................................................................... 56
   Instrumentation and Materials ...................................................................... 60
   Data Collection Procedures ......................................................................... 63
   Data Analysis ................................................................................................ 64
   Protection of Participants ............................................................................ 65
   Conclusion ..................................................................................................... 66

4. ANALYSIS OF THE DATA ............................................................................ 67
   Data Analysis ................................................................................................ 70
   Conclusion ..................................................................................................... 99

5. SUMMARY AND CONCLUSIONS .................................................................. 103
   Overview ...................................................................................................... 103
   Summary of Research .................................................................................. 103
   Summary of Findings ................................................................................... 105
   Discussion ..................................................................................................... 107
   Implications ................................................................................................. 109
   Areas for Future Research .......................................................................... 117
   Conclusion .................................................................................................... 118

xiv
APPENDICES .......................................................................................................................... 120

A: Teacher Survey ............................................................................................................. 121

B: Principal Survey ......................................................................................................... 125

C: Teacher Interview ........................................................................................................ 128

D: Principal Interview ...................................................................................................... 130

REFERENCES ....................................................................................................................... 132
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. School Demographics</td>
<td>57</td>
</tr>
<tr>
<td>2. Number of Students Studied</td>
<td>58</td>
</tr>
<tr>
<td>3. Number of Students Studied by Grade</td>
<td>58</td>
</tr>
<tr>
<td>4. Number of Students Studied by Ethnicity</td>
<td>59</td>
</tr>
<tr>
<td>5. Number of Students Studied by Language Proficiency</td>
<td>59</td>
</tr>
<tr>
<td>6. Mean Value of Test Scores for All Students</td>
<td>71</td>
</tr>
<tr>
<td>7. Paired Samples Test for All Students</td>
<td>72</td>
</tr>
<tr>
<td>8. Mean Value of African American Students</td>
<td>73</td>
</tr>
<tr>
<td>9. Paired Samples Test for African American Students</td>
<td>74</td>
</tr>
<tr>
<td>10. Mean Value of Asian Students</td>
<td>75</td>
</tr>
<tr>
<td>11. Paired Samples Test for Asian Students</td>
<td>76</td>
</tr>
<tr>
<td>12. Mean Value of Latino Students</td>
<td>77</td>
</tr>
<tr>
<td>13. Paired Samples Test for Latino Students</td>
<td>78</td>
</tr>
<tr>
<td>14. Mean Value of Non-EL Students</td>
<td>78</td>
</tr>
<tr>
<td>15. Paired Samples Test for Non-EL Students</td>
<td>79</td>
</tr>
<tr>
<td>16. Mean Value of EL Students</td>
<td>81</td>
</tr>
<tr>
<td>17. Paired Samples Test for EL Students</td>
<td>81</td>
</tr>
<tr>
<td>18. Mean Value of Second Grade Students</td>
<td>82</td>
</tr>
<tr>
<td>19. Paired Samples Test for Second Grade Students</td>
<td>83</td>
</tr>
</tbody>
</table>
20. Mean Value of Third Grade Students..............................................................84
21. Paired Samples Test for Third Grade Students .............................................85
22. Mean Value of Fourth Grade Students............................................................86
23. Paired Samples Test for Fourth Grade Students .............................................87
24. Mean Value of Fifth Grade Students..............................................................88
25. Paired Samples Test for Fifth Grade Students .................................................89
26. Teacher Survey Responses ............................................................................90
27. Significant Difference by Subgroups .............................................................100
Chapter 1
INTRODUCTION

The United States is a nation that strives to be exceptional and lead in a global economy unmatched by any other. The demand to be the best has resulted in the nation’s intense focus on quality education for all children, as they are the future of this country. Education is a powerful tool for reducing poverty and inequality along with setting a foundation for creating and sustaining economic growth and opportunities (The World Bank Group, 2011). It plays an important role in creating productive, efficient, pliable, and collaborative workers and critical thinkers as well as expanding a worker’s aptitude to successfully complete tasks or use industrious technologies (Decker, 1997; Radcliffe, 2009). According to the U.S. Department of Labor (1990), increases in educational opportunities were responsible for 20% growth in worker productivity in the United States in recent decades.

Considering the critical role of education, the report *A Nation at Risk* was essential in determining the nation’s state of education. In 1983, the Secretary of Education, T. H. Bell, created the National Commission on Excellence in Education to report on the quality of education in America. The report described main issues affecting American education and provided recommendations for educational improvement. Secretary Bell created the Commission out of his concern for increasing negative national perceptions of American education. According to the report, “…the educational
foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people” as the nation’s educational attainments were being matched and surpassed by others throughout the world (A Nation at Risk, 1983, p. 9).

Sixteen years after *A Nation at Risk* report, The Public School Accountability Act (PSAA) was passed in California in 1999 as the initial step toward developing a comprehensive system that holds students, schools, and districts responsible for improving student performance. There are three central components of PSAA: Academic Performance Index (API), Intermediate Intervention/Underperforming Schools Program (II/USP), and the Governor’s Performance Award (GPA). Schools are ranked by the Academic Performance Index (API), which measures academic performance and schools’ achievement growth ranging from 200-1000, with a statewide API goal of 800.

The II/USP financially supports low performing schools’ efforts to improve students’ academic outcomes with the condition that these schools agree to future sanctions if they do not improve. The GPA, also known as the High Achieving/Improving Schools Program (HA/ISP), financially rewards schools that meet their schoolwide API growth target, show comparable growth among all significant subgroups of students, and satisfy participation rates (EdSource, 2003). However, this component of PSAA is no longer offered as a result of little overall effect on student achievement in II/USP schools (California Department of Education, 2003). As part of the accountability act, students in grades 2-11 are tested through the Standardized Testing
and Reporting (STAR) program and in high school, students must also take the California High School Exit Exam (CAHSEE); both tests are aligned to academic content standards. Student achievement is based on test data, and graduation and attendance rates. The primary goal of PSAA is to raise student test scores and overall achievement in California’s low performing schools by helping schools improve and measure academic achievement for all students (Brodie, Crowley, & Kim, 2001; California Department of Education, 2011).

Nineteen years after A Nation at Risk, the Elementary and Secondary Education Act (ESEA) was reauthorized as the No Child Left Behind Act (NCLB) Act. NCLB was signed into law in 2002 as a result of Congress’s and President George Bush’s efforts to improve student achievement and educational opportunities for all children, regardless of ethnicity, income, or background (U.S. Department of Education, 2003). As stated by President Bush, “Too many of our neediest children are being left behind” (U.S. Department of Education, 2003, p. 7). NCLB’s “…mission is to ensure equal access to education and to promote educational excellence throughout the nation” (U.S. Department of Education, 2003, p. 3). The Act focuses on effective instruction and best practices that ensures all children read by third grade.

Education plays an enormous and vital role in society as well as the United States’ economy, thus warranting laws mandating schools to implement accountability measures in response to increasing concerns about American education. Districts and schools across the nation are pressured to produce academically proficient students who are well-
prepared to work, live, and contribute effectively upon graduation. Proficient student performance on high stakes standardized tests determines how academically qualified students are in their preparation for college or the workforce. Consequently, the increasing focus on preparing students for statewide testing is understandable as it has become imperative that students demonstrate and achieve academic proficiency on these high stakes state standardized tests.

**Impact of High Stakes Testing**

The prominence placed on high stakes testing in American education is the current threat amongst schools and districts nationwide (Mesler, 2008). In 2008, Mesler investigated the influence of high stakes testing on teaching methods, school curricula, and overall school management. More specifically, Mesler examined how schools use data produced by state tests to inform instructional practices. Using statewide testing data available on the North Carolina Public School database, Mesler identified different schools throughout the state that were succeeding and schools that were struggling to meet standardized test and growth expectations. Using data organized by grade, school, race, socioeconomic status, school size, teacher-to-student ratios, and overall rankings, she selected three different schools to study (*Honor School of Excellence, School of Excellence*, and *Priority School*) to distinguish between succeeding and struggling schools.
Mesler surveyed and interviewed three seventh grade English language arts teachers from identified schools to understand the impact of high stakes testing on schools. Results from Mesler’s study revealed that teachers at all three schools, whether Honor School of Excellence, School of Excellence, or Priority School, were all focused on either maintaining or improving standardized test scores. All three teachers interviewed made changes to their teaching methods. A Priority School teacher made changes consistent with several large studies examining the effects of high stakes testing in individual classrooms (Cimbricz, 2002; Marchant, 2004; Paris & Urdan, 2000) including using more skill- and test-oriented activities, and classroom activities that are more quantifiable so data on student achievement can be gathered. Based on Mesler’s interviews, teachers were focusing specifically on material covered by standardized tests rather than faithfully utilizing the district adopted curriculum and curricular materials. According to a School of Excellence teacher, only what gets tested gets taught, similar to practices of teachers at Priority schools and curriculum and curricular materials as secondary resource for increasing student achievement on standardized tests. Mesler concluded there were few qualitative differences between the three schools, whether they were struggling or succeeding to meet state standards.

High Stakes Testing and Staff Development

In addition to influencing teaching methods, school curricula, and overall school management as described in Mesler’s (2008) study, high stakes testing also impacts staff development. Berry, Turchi, and Johnson (2003) presented case study findings from a
cross state investigation of 24 schools in 12 districts with different accountability performance ratings from six states. The study documented and explained how teachers from schools with different performance ratings responded to accountability requirements and how their school districts responded to professional development likely to improve student achievement. According to their site visits, the researchers found that high stakes testing helped teachers in both low and high performing schools by focusing on academic standards and providing clarification on what needs to be taught and how. Teachers further supported high stakes testing, by uniformly reporting that professional development had improved and were more focused on “hands-on” learning about how to teach state standards.

**Benchmark Assessments**

Although educators initially focused their attention on annual state tests, they quickly realized annual test results come too late to identify students who are falling behind. “State tests can be powerful motivators, communicating expectations and focusing curriculum and instruction, but they rarely provide ongoing information that schools need to guide instructional programs and address learning problems of students who might otherwise be left behind” (Herman, 2005, p. 48). On the other hand, evidence from other assessments, particularly teachers’ classroom formative assessments, are more effective in guiding and informing instruction and could potentially lead to significant gains in student learning (Black & Wiliam, 1998).
In 2007, researchers Henderson, Petrosino, Guckenburg, and Hamilton, of Learning Innovations at WestEd (a REL-NEI partner organization), conducted a quasi-experimental study which investigated the impact of benchmark assessments on student math achievement in Massachusetts. Student test scores from 22 pilot program schools and 44 comparison schools were measured on the eighth-grade Massachusetts Comprehensive Assessment System (MCAS) math exam from 2001 to 2007. No statistically significant difference in test scores could be found between comparison groups. Petrosino and Henderson (2007) cautioned against drawing conclusions about benchmarks assessments based on their study, considering they had no control over all school variables pertaining to leadership, student motivation, teacher training, and how schools use the benchmark data. They believe more years of data collection were necessary to accurately gauge the effect of the benchmark assessment program on student scores.

Additionally, Blanc, Christman, Liu, Mitchell, and Travers (2010) conducted a multi-method study which examined the use of interim assessments in five elementary schools in the Philadelphia School District. The qualitative component of the study drew on data from fieldwork and case study which resulted in substantial evidence that interim assessments have the potential to contribute to instructional improvement. The quantitative component of the study showed greater achievement growth at schools where there was more evidence of fidelity to the curriculum and implementation of professional development regarding the use of student data. More importantly, the
research concluded that “…the success of even a well-designed system of interim assessments is dependent on the knowledge and skills of the school leaders and teachers who are responsible for bringing the system to life in schools” (Blanc et al., 2010, p. 222).

Currently, many districts and schools nationwide are administering periodic benchmark assessments to complement the end-of-year state testing and increase student learning as a response to accountability pressures to raise student achievement despite the lack of research base concerning its effectiveness (Henderson, Petrosino, Guckenb уч, & Hamilton, 2007). These assessments are used to gain critical information to guide instruction, monitor student learning, predict scores on future state tests, and identify students who are likely to score below proficient on state tests (Henderson et al., 2007). The question remains whether there is a significant difference between scores on the last benchmark assessment and state test scores subsequently taken at a later date? Moreover, how do benchmark assessments impact teaching practices of teachers and leadership practices of principals at Title I Program Improvement (PI) schools?

**Title I PI Schools**

Title I PI schools are low achieving with a high percentage of students from low income families. Title I of ESEA distributes funding to school districts which in turn distributes the money to low achieving schools that have significant enrollments of students who are from low income families. Title I is the oldest and largest federally
funded program in the nation in support of students at risk of failure or living at or near poverty. This policy, committed to closing the achievement gap between low income students and other students, was originally enacted in 1965 under ESEA but was rewritten in 1994 to improve essential goals for helping at-risk students. Title I schools that fail to make Adequate Yearly Program (AYP) for two consecutive years in the same content area such as English-language arts or math schoolwide or for any numerically significant subgroup, or on the same indicator (Academic Performance Index [API] or high school graduation rate) schoolwide are identified as PI schools.

In 2009-10 school year, NCSD selected 14 elementary and one middle school to pilot the CAB and participate in professional developments. The pilot did not start until February 2010. Currently, all schools with grades 2-8 are expected to give the first three benchmarks for math and ELA. Many schools in NCSD use benchmarks with the exception of a few specialized schools that may be therapeutic in nature or serve students who are enrolled in independent studies.

**Statement of the Problem**

Benchmark assessments are presumed to help increase students’ academic achievement on statewide tests despite the lack of research base concerning its effectiveness (Henderson et al., 2007). Up to the present, no study has been conducted to investigate the relationship between student scores on the final benchmark assessments and student scores on the CSTs in Title I Program Improvement elementary schools in a
northern California school district. Furthermore, no study exists on the influence of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals in these low achieving elementary schools.

**Research Questions**

This research will attempt to answer the following questions:

1. Is there a significant difference between students’ Curriculum Associates Benchmark (CAB) scores and students’ California Standards Test (CST) scores?
   a. Is there a significant difference between student scores on the final benchmark assessment in English Language Arts and student outcomes on the California Standards Test?
   b. Is there a significant difference between student scores on the final benchmark assessment in mathematics and student outcomes on the California Standards Test?
   c. Are student scores on the final benchmark assessments in English Language Arts and student outcomes on the California Standards Test significantly different for student subgroups (EL, Non-EL, African American, Asian, and Latino)?
   d. Are student scores on the final benchmark assessments in mathematics and student outcomes on the California Standards Test significantly different for student subgroups (EL, Non-EL, African American, Asian, and Latino)?
Test significantly different for student subgroups (EL, Non-EL, African American, Asian, and Latino)?

e. Are student scores on the final benchmark assessments in English Language Arts and student outcomes on the California Standards Test significantly different by grade levels (grades 2, 3, 4, and 5)?

f. Are student scores on the final benchmark assessments in mathematics and student outcomes on the California Standards Test significantly different by grade levels (grades 2, 3, 4, and 5)?

2. In what ways have benchmark assessments influenced teaching practices of teachers in elementary Year 5 Title I Program Improvement schools?

a. What modifications, if any, have teachers made in their instructional approach?

b. How have benchmark assessments impacted the ways in which teachers plan and organize their lessons?

c. How have benchmark assessments changed the ways in which teachers work with specific groups of students?

d. What influence have benchmark assessments had on teacher practices regarding evaluation of student achievement?

3. In what ways have benchmark assessments influenced the instructional leadership practices of principals in elementary Year 5 Title I Program Improvement schools?
a. How have benchmark assessments influenced principals to modify their leadership approaches?

b. How have benchmark assessments impacted the ways in which principals monitor instruction?

c. How have benchmark assessments changed the ways in which principals work with teachers?

**Theoretical Base**

This study will apply the concept of Bolman and Deal’s (2008) four organizational frames (structural, human resource, political, and symbolic) to the subject of benchmark testing. The influence of benchmark testing on teaching practices of teachers and instructional leadership practices of principals will be examined with reference to these four frames. Bolman and Deal (2008) believe viewing issues from multiple perspectives (reframing) is necessary to gain clarity, balance, generate innovations, and discover critical strategies that make a difference.

The *structural frame* focuses on the architecture of the organization, the design of its units and subunits, its rules and roles, its goals and policies; this frame emphasizes the importance in putting people in the right roles and relationships (Bolman & Deal, 2008). The *human resource frame* focuses on understanding people, their strengths and weaknesses, reason and emotion, desires and fears, and stresses dealing with issues by changing people (via training, rotation, promotion, or dismissal) (Bolman & Deal, 2008).
The political frame views organizations as competitive and resource deficient, struggling for power and advantages (Bolman & Deal, 2008). The symbolic frame addresses issues of meaning and faith, with rituals, ceremonies, stories, plays, and cultures as central organizational life (Bolman & Deal, 2008).

**Operational Definitions**

In addition to Bolman and Deal’s (2008) four organizational frames, the following terms are used throughout this paper:

*Accountability* focuses on how much schools are improving in year-to-year academic growth (http://www.cde.ca.gov/ta/ac/pa/cefpsaa.asp).

*Academic Performance Index (API)* is the cornerstone of the state’s academic accountability requirements. Its purpose is to measure the academic performance and growth of schools. The API is based on the results of statewide tests in grades two through twelve. (http://www.cde.ca.gov/ta/ac/pa/cefpsaa.asp)

*Adequate Yearly Progress (AYP)* is a statewide accountability system mandated by the No Child Left Behind Act of 2001 which requires each state to ensure that all schools and districts make Adequate Yearly Progress (http://www.cde.ca.gov/ta/ac/ay/).

*Assessment* is a process that gathers information about and measures what students know and are able to do.

*Benchmark Assessments* are tests that are administered periodically throughout the year to obtain immediate formative feedback on how students are learning.
*California Standards Tests (CSTs)* is a major component of the Standardized Testing and Reporting (STAR) program that measure students’ progress toward achieving California’s state-adopted academic content standards in English-language arts (ELA), mathematics, science, and history-social science.

*Elementary and Secondary Education Act (ESEA)* is a 1965 U.S. federal statute that funds elementary and secondary education with emphasis on equal access, high expectations, and accountability.

*Formative Assessments (Assessment for Learning)* measures student learning as it is happening in order to modify instruction and learning activities to improve student achievement.

*High Stakes Testing* is high pressure and involves important consequences.

*No Child Left Behind Act of 2001 (NCLB)* is a reauthorization of the Elementary and Secondary Education Act (ESEA) focused on all students, particularly those who are historically underserved, poor, and culturally and linguistically diverse students, in order to achieve academic proficiency.

*Numerically Significant:* 100 or more students with valid STAR test scores or 50 or more students enrolled with valid test scores who make up at least 15 percent of the total valid test scores.

*Public School Accountability Act (PSAA):* A 1999 comprehensive system that holds students, schools, and districts accountable for improving student achievement.
**Subgroup** is the number of students with valid STAR scores that contribute to the subgroup's Growth API. A numerically significant subgroup for the API is a group of 100 or more students with valid STAR scores or 50 or more students with valid STAR scores who make up at least 15 percent of the total valid STAR scores. A subgroup must be numerically significant in both the base year and growth year in an API reporting cycle to have subgroup growth and target information. Subgroups include: African American (not of Hispanic origin), American Indian or Alaska Native, Asian, Filipino, Hispanic or Latino, Pacific Islander, and White (not of Hispanic origin).

([http://www.cde.ca.gov/ta/ac/ap/glossary09e.asp](http://www.cde.ca.gov/ta/ac/ap/glossary09e.asp))

**Summative Assessments (Assessment of Learning)** are cumulative tests used to evaluate and measure long term student growth after a period of instruction and generally given at the end of a course ([http://www.learnnc.org/lp/pages/5233](http://www.learnnc.org/lp/pages/5233)).

**Significance of the Study**

Given the national priority on high stakes testing, many schools and districts in the nation are using benchmark assessments to improve teaching and learning. Studying the relationship between benchmark assessments and state tests and the influence of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals will benefit many stakeholders: students, teachers, principals, schools, and school districts. First, by investigating if there is a significant difference between student scores on benchmark assessments and student scores on state tests, all
stakeholders can better plan ways to increase student achievement on high stakes tests. For example, if there is a strong and positive relationship between the two, school districts may be willing to invest more time and money on providing professional developments to teachers and principals to strategically use benchmark assessments to plan instruction to increase student achievement. Additionally, principals may require teachers to incorporate benchmark assessments into their daily routine or to plan instruction to meet specific student needs based on data from previously taken benchmark assessments. Also, teachers may use benchmark assessments as a roadmap to guide what students are expected to know and to plan instruction accordingly with regards to curriculum and curricular materials. Ultimately, students will know that doing well on benchmark assessments relates to them doing well on the state tests, therefore, lowering their affective filter. In the same token, students who struggle to do well on benchmark assessments are given the opportunity to improve when teachers strategically focus on their areas of needs.

**Limitations of the Study**

This study is limited in scope by the selection of participants and small sample size from two public elementary schools in one of the top ten largest school districts in California along with findings limited to benchmark assessments created by one company. However, as Title I Program Improvement schools, the schools in this study, in many ways, mirror the demographics, experiences, and concerns of others across the
nation. The study is further limited by the few numbers of principals, one specifically, who chose to participate in the survey and declined to be interviewed. It is hard to generalize to the population due to the limited size. Furthermore, the CAB cannot be generalized to other benchmarks. Moreover, the study is limited to only students with both CAB and CST scores.

Like any survey- and interview-based study, this research design assumes that teachers and administrators will understand the survey questions and self-report accurately. The study of responses from educators working in Title I schools may not be reflective of responses from those working in non-Title I schools.

Conclusion

Many schools and districts nationwide are increasingly administering periodic “benchmark” assessments to complement the end-of-year high stakes state test. However, the research base concerning its effectiveness is limited (Henderson et al., 2007). Benchmark assessments serve many uses: gain critical information to guide instruction, monitor student learning, predict scores on future state tests, and identify students who are likely to score below proficient on state tests (Henderson et al., 2007). Investigation is needed to determine if there is a significant difference between scores on the last benchmark assessment and scores on state tests (Henderson et al., 2007). The answer to this question may inform current practice with benchmark assessments in Title I Program Improvement schools, and ultimately improve teaching and learning.
Chapter Two will review the literature on high stakes testing and benchmark assessments. Chapter Three will specify the research methods used to determine the relationship between benchmark assessments and statewide assessments and the influence of benchmark assessments on teaching of teachers and instructional leadership practices of principals. After completion of the research, Chapters Four and Five will be utilized to discuss the findings of this study, and the implications for Title I Program Improvement elementary instruction, leadership, and policy.
This review of the literature will examine NCLB, high stakes testing, benchmark assessments, data-drive decision making, and instructional leadership. In addition, the research on Curriculum Associates Benchmark (CAB) will be presented. Research on this topic has tended to focus on the impact of high stakes testing and accountability policies. There is a lack of information specifically on benchmark testing. This review of the literature will conclude by arguing the need to examine the relationship between benchmark assessments and statewide assessments, and the influence of benchmark testing on teaching practices of teachers and instructional leadership practices of principals.

**No Child Left Behind and High Stakes Testing**

The American federal legislation education reform, No Child Left Behind (NCLB) Act of 2001, established rigorous expectations that all students receive high quality instruction and achieve academic proficiency (Muller & Schiller, 2000; Ormrod, 2006; U.S. Department of Education, 2002). The passage of NCLB also resulted in an influx of interest in using benchmark assessments to improve student learning (Herman et al., 2010). NCLB accountability policy focuses explicitly on educational challenges that call for closing the achievement gap, particularly of disadvantaged, poor, and minority
children who have historically been underserved (Bancroft, 2009; Hong & Youngs, 2008; Noche, 2007). As part of the 2002 reauthorization of Title I of the Elementary and Secondary Education Act (ESEA), NCLB mandates annual testing in reading, writing, and mathematics for students in grades 3-8 and once during grades 10 through 12 in order to measure and monitor academic proficiency for all ethnic groups in schools and districts in the United States. This mandate was created in order to prevent students from moving up in grade level year after year with passing grades only to demonstrate abysmal skills upon graduating, which has frequently been the case for urban city children (Hong & Youngs, 2008; Noche, 2007).

These annual tests are considered to be high stakes as they hold serious consequences related to school funding, student graduation, scholarship opportunities, and teacher and school reviews (American Educational Research Association, 2000; Amrein-Beardsley, 2009). Both students and educators in Title I schools and districts that fail to make Adequate Yearly Progress (AYP) face serious consequences (Hong & Youngs, 2008; Linn & Miller, 2005; Marchant, 2004; Ormrod, 2006). High stakes testing is a method used to judge schools and teachers by the percentage of students attaining academic proficiency (Ross, 2008). High stakes tests gauge how well students have attained grade level standards and are used for examining and improving instructional practices in addition to student performance outcomes (Bancroft, 2009). According to Sadker and Zittleman (2004), “Standards are like guiding stars that enable us to navigate a successful academic voyage” (p. 740). Bancroft (2009) stated that
“because standardized testing is now the predominant method for capturing the value of a school’s academic stature, an examination of the school’s test scores is requisite for understanding the struggles of schools that serve poor communities with correlating poor test score results” (p. 53).

Consequently, the current focus of American education is on constructing common standards and benchmarks as well as developing and using more uniform standards-based tests for all students (Wieczorek, 2008). These practices would ultimately improve academic performance for low-income and minority students resulting in decreased inequality in education (O-Day & Smith, 1993). Standards-based reform supports educating all children from all social class and racial backgrounds as it pressures teachers to critically reflect on and modify their instructional strategies to increase student test scores (Cimbricz, 2002; Marchant, 2004; Paris & Urdan, 2000, Ravitch & Hirsch, 1995). Mesler (2008), Paige (2001), Grissmer, Flanagan, Kawata, and Williamson (2000), and others advocate NCLB’s high stakes testing and similar state policies guiding teachers, schools, and districts’ commitment to high standards for low socioeconomic and minority students achieve high levels of success.

**Effects of High Stakes Testing**

High stakes testing is an extremely controversial issue. The implausible ramifications of high stakes testing and accountability policies have been extensively documented (Darling-Hammond, 2004a; Kozol, 2005; Noguera, 2003; Reddell, 2010; Rose, 1996; Shepard, 2000; Valenzuela, 2004). According to Hong and Youngs (2008),
high stakes testing does not support student achievement. For example, a 2001 national study by Jacob, concluded that students in the bottom fifth of their grade at schools that required high stakes testing were 25% more likely to drop out of high school than their counterparts who were not required to take the test. Additionally, students who were retained from progressing to the next grade level as a result of low test scores on high stakes tests were more likely to drop out of school (Marchant, 2004). Some studies also indicate deleterious effects are evident for curriculum, instruction, and student dropout rate (Hong & Youngs, 2008).

For example, Lowry (2010) examined teachers’ perceptions of NCLB and four variables: instructional practices, testing practices, professional development, and school climate within a high stakes testing environment and a non-high stakes testing environment, in which she collected data from 102 teacher questionnaires. The findings for this quantitative study indicated that both teachers in high stakes testing environments and teachers in non-high stakes testing environments spent a significant amount of time on test preparation activities. However, high stakes teachers tended to spend more time on test preparation activities than non-high stakes teachers. Furthermore, the study also indicated that professional development significantly affects the climate of a school and teachers who are more positive toward professional development are more likely to identify the school climate as focused on test preparation.

Additionally, in an in-depth qualitative case study by Tingey (2009), which investigated the impact of high stakes testing under the NCLB Act on school culture,
individual and focus group interviews were conducted with first grade through sixth grade teachers and principals from two Nebo School District’s schools located in Utah. In general, teachers and principals in this study reported that high stakes testing negatively impacted student and teacher motivation, teaching and learning, and curriculum with the exception that they did view testing data important in providing some accountability as well as some useful information about teaching and learning (Tingey, 2009).

On the other hand, some researchers have provided evidence that high stakes testing has resulted in increased student achievement (Braun, 2004; Carnoy & Loeb, 2002; Rosenshine, 2003; Scheurich, Skrla, & Johnson, 2000; Schiller & Muller, 2000). Hong and Youngs (2008) reported that students’ performance on both state and district tests rose in Texas and Chicago in the 1990s after the implementation of high stakes testing and accountability policies. However, upon re-examination of the phenomenal gains in student achievement, researchers (Amrein-Beardsley & Berliner, 2000; Haney, 2000) found the percentage of students excluded from participating in the test increased whereas the remainder of the nation’s exclusion rate declined concurrently.

Low Performing Schools

Data from interviews and observations regarding how schools respond to high stakes accountability policies indicate the significance of school context (Cole, 1996; Diamond & Spillane, 2000; Hong & Youngs, 2008; Wetsch, 1998). For example, low performing schools are narrowly focused on complying with policy demands and
improving non-proficient performance for tested students in reading and mathematics (Diamond & Spillane, 2000). They are more predisposed to focus mainly on the general overall test results (the need to improve reading and math), rather than concentrate on identifying systematic strategy for using test results as tools for instructional alteration (Diamond & Spillane, 2000). Students at low performing schools who are on the verge of achieving proficiency are generally provided extra assistance such as after school tutoring to help them reach proficiency level (Diamond & Spillane, 2000). Furthermore, some low-performing schools are devoted to impressing district personnel by attending to superficial transformations such as classroom management, décor, and student work display (Diamond & Spillane, 2000).

In one particular study, upon reviewing a North Carolina statewide testing data, Mezler (2008) selected and compared three different successful and struggling schools to meet standardized test and growth expectations to determine the influence of high stakes testing on teaching and organization. Results from Mesler’s study indicated that the struggling school focused on improving test scores so they could get out of priority status. The school made several significant changes consistent with several large scale studies on the effects of high stakes testing in individual classrooms (Cimbricz, 2002; Marchant, 2004; Paris & Urdan, 2000), including block scheduling so teachers would have more time to concentrate on teaching test specific topics in addition to providing skill and test related activities; on top of benchmark testing.
**High Performing Schools**

On the contrary, high performing schools are inclined to use the full range of test score data and focus on informing and identifying specific instructional needs of particular students and channel their energy to increasing performance for all students in all grade levels in all subject areas (Diamond & Spillane, 2000). High performing schools use test results to identify high and low performing groups of students as well as diagnose instructional effectiveness and adopt interventions for all students (Diamond & Spillane, 2000).

In a 5-year National Research Center on English Learning and Achievement (CELA) study of literacy skills of 25 middle and high schools in California, Florida, Texas, and New York, which explored “typical” and “high” state test performers, it was concluded that high performing schools embedded test preparation into the curriculum and literacy lessons were rich and varied (Langer & Applebee, 2000). Further, an in-depth survey of 257 high poverty elementary schools and 5,500 principals and teachers in California which compared low, middle, and high performing schools, revealed that high performing schools placed priority on student achievement, used assessment data to inform instructional decisions, as well as allocated resources to support student learning (Barth, 2006).

**Opponents of High Stakes Testing**

Opponents of high stakes testing believe such policies encourage “teaching to the test” or narrowing the curriculum to only what is tested (Amrein-Beardsley et al., 2010;
Barth, 2006; Berliner, 2009; Coffey, 2009; Jones & Egley, 2007; Lowrie, 2009; Marchant, 2004; McNeil, 2000; Mesler, 2000 & 2008; Shepard, 2000; Thompson, 2001). Additionally, high stakes testing focus instructional priorities on reading and mathematics; therefore, diminishing the importance of other subject areas (Cimbricz, 2002; Diamond & Spillane, 2000; Jacobs, 2002; Stecher & Barron, 2001). Furthermore, high stakes testing ignore skills that are critical to the real world (Darling-Hammond, 2004a).

Although teachers emphasize testing will distort instruction, few believe it affects their own teaching (Barth, 2006). More specifically, despite that 79% of teachers worry testing will lead to “teaching to the test” as opposed to a focus on “real learning,” 73% believe testing has not taken away from “real learning” in their classrooms (Barth, 2006).

According to Barth (2006), 71% of school districts in the nation reported increasing time for reading and mathematics to accommodate student proficiency performance on the state tests. Moreover, much of classroom time is devoted to test preparation instead of student learning, along with reducing learning to memorization and fact recall (Davis et al., 2002). Additionally, accountability policies incidentally encourage incentives that marginalize low performing students (Clotfelter & Ladd, 1996; Diamond & Spillane, 2000; Ross, 2008) as schools struggle to maximize the percentage of proficient students (Diamond & Spillane, 2000; Ross, 2008). Consequently, resources are allocated to helping students who are on the verge of proficiency while limited assistance is left over for improving scores for the lowest performing students who will
unlikely contribute to increasing the overall school performance (Diamond & Spillane, 2000; Reback, 2006; Ross, 2008). Such practices result in a major segment of the student population being left behind (Stiggins, 2002).

In a worst case scenario as described in Amrein-Beardsley, Berliner, and Rideau’s 2010 study, the pressure for students to perform well has resulted in educators cheating in various ways. The study investigated the types of, degrees to which, a sample of Arizona teachers were aware of or had engaged in test-related cheating practices due to high stakes testing policies of NCLB (Amrein-Beardsley et al. 2010). Valid survey responses for the study were obtained from a little over 3000 teachers in addition to a small convenience sample of teacher interview and a focus group. Data from the survey and interview revealed educators cheated in various ways. Cheating may be willful and premeditated as in the case of erasing and changing students’ test answers; giving students the correct answers during class; not testing academically weak students; or changing students’ identification number so low achieving students test scores are eliminated from the results. Cheating may be casually, as in the case of encouraging students to redo problems as students are testing; allowing students to use cheat sheets; allowing students to have access to classroom resources; or providing students with hints or tips for answering questions. Cheating may also be indifferent, reckless, or negligent as in the case of using knowledge about the test to give students an edge so they may do well.
Proponents of High Stakes Testing

NCLB equalizes educational opportunities for all students (Hong & Young, 2008) and pressures teachers to confront ineffective practices and push them to higher levels of performance (Marshall, 2008). Advocates believe such policies provide motivations for school improvement (Diamond & Spillane, 2000). By producing access to more objective information on student performance, NCLB is reducing inequality, increasing student motivation, and improving student outcomes (Coleman, 1997; Shouse, 1997; Muller & Schiller, 2000). NCLB has also increased numbers of students meeting minimum competency requirements according to Chicago data on the implementation of high stakes testing (Roderick et al., 1999). Additionally, it has produced greater student achievement gains based on research by Carnoy and Loeb (2002) along with promoting student learning (Porter, 2000). Moreover, according to investigations by Raymond and Hanushek (2003) and Winter (2003), students who test in states under high stakes accountability score higher in math on the National Assessment of Educational Progress (NAEP) than students who test in states without high stakes accountability. Likewise, accountability systems also resulted in overall improvements in student performance on NAEP reading and mathematics tests (Hanushek & Raymond, 2004).

The Council of Chief State School Officers (CCSSO) has been using state assessment scores and other data to analyze achievement trends to track and report on student achievement results (Blank, 2011). The CCSSO analysis of student achievement trends at the state-aggregate level show that during the period of time when grade-level
testing was fully implemented in reading and math (as required under NCLB), most states made significant gains in the student achievement performance of all students, including economically disadvantaged students for both subjects.

**Instructional Leadership**

Instructional leadership is critical to teaching and learning. It is collaborative versus top-down authoritarian and promotes improved student performance, learning, and growth. In general, principals are expected to be instructional leaders who are more skilled and knowledgeable than anyone else in the school and could provide guidance to teachers on how to teach students effectively (Hoerr, 2007/2008). Furthermore, instructional leaders are expected to communicate clear educational vision and offer direction and expertise that ensure all students learn (Hoerr, 2007/2008). According to Reading First Notebook (U.S. Department of Education, 2005), there are five key elements of instructional leadership: Prioritization, Scientifically based reading research (SBRR), Focus on alignment of curriculum, instruction, assessment, and standards; Data analysis, and Culture of continuous learning for adults. With reference to data analysis, instructional leaders must use multiple sources of information to assess student performance (NAESP, 2001). Decisions of all stakeholders (central office staff, principal, and teachers) must be based on appropriate data. Central office staff can use school data to guide principals in making policy and curriculum decisions and principals
can use classroom data to help teachers in grouping students for intervention (U.S. Department of Education, 2005).

Data-Driven Decision Making

Data plays an important role in improving education (teaching and learning) for children in public schools across the United States. Additionally, policy makers use data to evaluate programs and research to design and support programs and interventions in efforts to improve and increase academic proficiency level (Springboard Schools, 2006). Moreover, as a result of NCLB, data are now being disaggregated to track student achievement for different groups of students. Researchers like Thorn (2002) and Wayman, Midgley, and Springfield (2005) amongst many others, believe that teachers and principals must have access to formative or diagnostic summative data that provides information regarding student progress and achievement in order to improve curricular and school.

Currently, data is being used nationwide to support numerous strategies and instructional decision-making that positively affects children (Springboard Schools, 2006). Based on review of the literature, the demand for data that is linked to improved student outcomes is high. Although numerous descriptive literatures exist, few address the effects of data-driven decision making on student outcomes. Data is abundant at the state, district, and school levels; however there are many unanswered questions about
how to interpret and use data to inform decisions about student achievement and other educational outcomes (Marsh, Pane, & Hamilton, 2006).

**Assessments**

Although high stakes testing is indubitably controversial, school districts and administrators throughout the U.S. are nevertheless pressured to increase student achievement. Therefore, they are making alignment of instruction to high stakes standardized testing a major focus of student learning (Louis, Febey, & Schroeder, 2005). By the same token, they are also using benchmark assessments as an educational improvement strategy (Goren, 2010) that assists in monitoring student progress toward grade level standards mastery throughout the year prior to taking the exam at the end of the year to ensure a greater level of success (Herman & Baker, 2005; Olson, 2005a; Swanson & Collins, 1991).

**Assessment Data**

Assessment in education is critical to teaching and learning (CTB/McGraw-Hill, 2008). As stated by Salvia and Ysseldyke (1995), assessment is “…the process of collecting data for the purpose of 1) specifying and verifying problems and 2) making decisions about students” (p. 3). Data provides knowledge and insight about student learning that allows teachers to adjust and transform instruction needed to assist students’ academic success (Technology Alliance, 2005). Using data to inform and redesign instruction increases student achievement (Fournier et al., 2009; Matthews et al., 2007).
Assessments inform instruction, not merely tell educators how students are performing and achieving (Sibley, Biwer, & Hesch, 2001). According to Lowrie and Diezmann (2009), assessments of student performance are increasingly influencing and possibly driving instructional practices in most schools throughout the nation.

Assessments provide valuable, imperative data to schools, school districts, and states concerning student performance and achievement that allow them to make academic achievement decisions regarding individual students, groups of students, instructional programs, resource allocation, and more (Chappuis, Chappuis, & Stiggins, 2009). The power of assessment to improve instruction and learning is greatest when the assessment occurs naturally within the learning environment and feedback is immediate (CTB/McGraw-Hill, 2008; Fournier, 2009).

According to CTB/McGraw-Hill (2008), multiple assessment data allow educators to realize many objectives: monitor educational systems for public accountability; provide information to enhance instructional practices; evaluate effectiveness of instructional practices; measure student achievement and skills mastery; and provide data and tools to improve student progress. In short, test scores are proxies for gauging academic knowledge (CTB/McGraw-Hill, 2008).

**Summative Assessment**

As previously noted, no single assessment can provide complete information regarding student progress (CTB/McGraw-Hill, 2008). A combination of assessments is necessary to determine whether all educational goals are being met (CTB/McGraw-Hill,
Assessments can be used summatively to evaluate what content standards students have learned by the conclusion of instruction (Garrison & Ehringhaus, 2007; Perie et al., 2007). Summative assessments are cumulative evaluations (Coffey, 2009; McAlpine, 2002) that sum up what students have learned over time (Arter, 2010; CTB/McGraw-Hill, 2008) and provide the overall level of students’ performance on annual learning standards (Chappuis et al., 2009; Herman et al., 2010). Results from summative assessments can shape how teachers organize their curricula or transform their instructional practices (Coffey, 2009).

The most common types of summative assessments include: state-mandated assessments, district benchmark or interim assessments, end-of-unit or chapter tests, end-of-term or semester exams. Summative “high stakes” end-of-the-year state assessments are powerful motivators that communicate and focus curriculum and instruction in the upcoming year (Chappuis et al., 2009). Additionally, these assessments also indirectly increase student achievement (Black, 1988; Popham, 2003; Stiggins, 2002) although they seldom provide ongoing information that are needed in schools to guide instructional programs and address student learning needs (Herman, 2005).

NCLB mandates that all students are tested annually in reading and mathematics. This nationwide intensification of testing and accountability in reading and mathematics has caused many school districts in the nation to engage in “high stakes” assessments (Noche, 2007; Sawchuk, 2009; Sibley, 2001). Despite this fact, there is evidence that more students are struggling with curriculum skills now as the percentage of students
placed in learning disability programs across the nation increased from 2.3% to 5.2% from 1978 to 1993 (Shapiro, 1996). According to Stiggins (2002), “Student achievement suffers because these once-a-year test are incapable of providing teachers with the moment-to-moment and day-to-day information about student achievement that they need to make crucial instructional decisions” (2002, p. 759). Summative assessments are “…sometimes referred to as assessment of learning, which documents how much learning has occurred at a point in time; its purpose is to measure the level of student, school, or program success” (Chappuis & Chappuis, 2007, p. 15).

**Benchmark Assessment**

Formative assessments, also known as interim, diagnostic, predictive, or benchmark assessments, are “…embedded in ongoing classroom instruction to inform immediate teaching and learning goals” (Herman et al., 2010, p. 2). In this study, the researcher will use the term “benchmark” for consistency purposes throughout this paper as it is considered a facet of formative assessment (Smith & Shrago, 2006). Benchmark assessments are meant to improve teaching and learning in important ways. First, they provide quick, reliable feedback on how students are progressing toward state standards. Second, they provide timely information about the strengths and weaknesses of classes and grade levels on essential standards for each content area. More specifically, benchmark assessments provide immediate, ongoing targeted information pertaining to students’ academic strengths and weaknesses that teachers may employ to increase
student learning (Chappuis et al., 2009; Coffey, 2009; CTB/McGraw-Hill, 2008, Noche, 2007; Olson, 2007; Sawchuk, 2009).

Benchmark assessments can be used formatively to inform upcoming instruction and are given at the beginning or middle of instruction to provide teachers information about what students have learned or have yet to learn at that point in time (Chappuis et al., 2009; Perie et al., 2007). Based on Black and William (1998), formative is defined as “encompassing all those activities undertaken by teachers, and or/by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (p. 7-8). In addition, Black and Wiliam claim that formative assessment is “…when anyone is trying to learn, feedback about the effort has three elements: recognition of the desired goal, evidence about present position, and some understanding of a way to close the gap between the two…” (1998, p. 8).

Benchmark assessments track student progress (Marshall, 2008) and is intended to promote student learning (Black et al., 2002). It provides a check into student progress and insights about student performance and achievement and serves as models for transforming classroom practices, resulting in improved student learning (CTB/McGraw-Hill, 2008; Davis et al., 2002; Herman, 2005; Larter & Sullivan, 1993, Marshall, 2008; Olson, 2005 & 2007, Regional Educational Laboratory, 2007). These assessments provide students multiple opportunities to display their ability to apply knowledge and comprehension throughout the year (Chappuis et al., 2009). According to the Wisconsin Department of Public Instruction, “Benchmark assessments occur within, between, and
among instructional units” (2007, p. 21). The purpose of benchmark testing is to provide all stakeholders (teachers, students, and administrators) academic information to track student progress toward achieving standards mastery and essential diagnostic feedback to guide instruction and improve student learning (Fournier et al., 2009; Herman, 2005; Herman & Baker, 2005; Noche, 2007).

Benchmark data are “real-time data” that gauges what students are learning and what teachers are currently teaching over the course of the year (Matthews et al., 2007; Sawchuk, 2009). When benchmark results are provided immediately, teachers can quickly use the feedback on how students are meeting academic standards to modify their instruction to better meet their students’ individual learning needs (Coffey, 2009; Fournier et al., 2009; Henderson, 2008; Matthews, 2007) in the form of small, focused intervention groups for students not meeting standard proficiency (Halverson, 2010; Marshall, 2008). Benchmark assessments that are aligned to state content standards, “…provide a continuous, comprehensive stream of information to plan and guide instruction” (Herman et al., 2010, p. 9), and provide intervention information to guide teaching and learning have the ability to guide curriculum needs and professional development of teachers as well as provide research-based predictions to student performance on state summative tests (Coffey, 2009; CTB/McGraw-Hill, 2008; Henderson, 2008).

**Predictive validity of benchmark testing.** As indicated by Southwest Regional Educational Laboratory (2007), previous research studies are narrowly focused, both in

Similar studies were analyzed in mathematics. Clarke and Shinn (2004) scrutinized the predictive validity of four curriculum-based measures in predicting first grade student performance on three distinct criterion measures in a school district in the Pacific Northwest. VanDerHeyden, Witt, Naquin, and Noell (2001) also included math outcomes in their review of the predictive validity of kindergarten student readiness in Louisiana.

Each study mentioned above focused on the predictive validity of the given benchmark assessment or state-mandated test. These studies generally showed that benchmark assessments have the potential to predict outcomes but there was great
variability in the extent of the existing relationships. On the contrary, according to Regional Educational Laboratory (2007), benchmark assessments generally lack predictive validity to student performance on state tests. Learning Innovations researchers Petrosino and Henderson cautioned against drawing conclusions about the lack of predictive validity of benchmark assessments due to many influencing variables for which schools cannot control, including leadership, student motivation, teacher training, and how schools use benchmark data (Henderson, Petrosino, Guckenburg, & Hamilton, 2008).

Given the primary focus of benchmark testing is used as a predictive tool of state tests, benchmark tests must adhere as closely as possible to the state test blueprints in terms of content, rigor, number of items per topic, cognitive demands, and format of test items as well as cover content targeted by the state test in each instructional period in order to maximize correlations with the state test (Marshall, 2008; Niemi et al., 2007). With that being said, according to a Mid-Atlantic Region report which examined the quality of predictive validity data for a selection of benchmark assessments identified by state and district personnel, evidence is generally lacking of their predictive validity with respect to state assessment tests (Brown & Coughlin, 2007). Brown and Coughlin (2007) investigated the relationship between district and state test scores as well as between performance on district-administered benchmark assessments and proficiency levels on state assessments. The study determined that while commonly used benchmark assessments in the Mid-Atlantic Region jurisdictions possessed strong characteristics,
their predictive validity with respect to the state or summative assessments was generally lacking.

The Mid-Atlantic Region reviewed four benchmark assessments: Northwest Evaluation Association’s Measures of Academic Progress (MAP; Northwest Evaluation Association, 2003), Renaissance Learning’s STAR Math/STAR Reading (Renaissance Learning, 2001a, 2002), Study Island’s Study Island (Study Island, 2006), and CTB/McGraw-Hill’s TerraNova (CTB/McGraw-Hill, 2001). Findings from the study documented degrees of criterion validity for three of the benchmark assessments (STAR, MAP, and TerraNova), but only one was found to be a predictive study that demonstrated strong evidence of predictive validity (TerraNova).

The multiple uses of benchmark assessments. In addition to predicting (TerraNova) student academic success on end-of-the-year high stakes state tests (Arter, 2010; Black & Wiliam, 1998), there are other uses for benchmark assessments, including evaluation of programs and curricula implementation, feedback for transforming instructional practices, and decision-making at the classroom, school, and district levels (Herman et al., 2010; Perie et al., 2007). Benchmark assessments are increasingly being utilized to monitor teaching and learning in efforts to close the achievement gaps (Bancroft, 2009; Blanc et al., 2010) as they have the potential to produce significant gains in student learning, achievement, and motivation on high stakes state tests (Arter, 2010; Black & Wiliam, 1998).
Benchmark assessments allow students to show standard mastery of what they have been taught in previous weeks or months and teachers to focus their instruction on what students still struggle to learn at the moment (Chappuis et al., 2009). These assessments systematically tap into students’ ability to apply knowledge and comprehension of what has been taught over the course of several weeks (Marshall, 2008) and give teachers data for diagnosing which instructional practices and learning objectives need improvement (Sawchuk, 2009). Benchmark assessments are typically given periodically three times a year to as often as once a month; provide timely and effective feedback to teachers and students; focused on reading and mathematical skills; reflect state or district academic content standards; and measure student progress on curriculum and state exams (Coffey, 2009; Herman et al., 2010; Olson, 2005; Pijanowski, 2008). More importantly, benchmark assessments have the power to help build student confidence and reduce stress level prior to taking the state test near the end of the year (Marshall, 2008).

**Research on how benchmark assessments are used.** Limited research exists on how benchmark assessments are used by teachers in the classroom, by principals, and by districts, as well as the impact of benchmark assessments on student performance. Studies that surveyed and interviewed teachers about their use of benchmark test results in instruction reported test results helped them monitor student progress and identify students’ skill gaps that led them to modify the curriculum and their instruction (Christman et al., 2009; Clune & White, 2008; Stecher et al., 2008). Based on a 2006
study of school districts in California, Georgia, and Pennsylvania, Stecher, Epstein, Hamilton, Marsh, Robyn, McCombs, Russell, and Naftel (2008) found that 53% to 73% of the elementary and middle school mathematics used benchmark assessments. Seventy-five percent of the teachers in this study confirmed that the benchmark test results helped them identify and modify gaps in curriculum and instruction (Stecher et al., 2008).

Similarly, 86% of teachers in Province, Rhode Island reported modifying their instruction as a result of benchmark assessments (Clune & White, 2008). They reviewed problems from the assessments, aligned their instruction with the benchmark assessments, paid more attention to weak skills, focused on content that would be covered during each benchmark assessment period (Clune & White, 2008), or retaught content and skills assessed on benchmark assessments (Christman et al., 2009). According to Nabors Olah, Lawrence, and Riggan (2010), who studied the use of benchmark assessments by teachers in Philadelphia, teachers used benchmark data to identify weaknesses in student performance in particular skills to reflect on their teaching practices. Furthermore, they used benchmark results to plan their instruction to focus specifically on areas students performed poorly on. Unfortunately, none of the aforementioned studies examined how teachers analyzed and used the assessment data to inform their classroom practice.

As much as there is limited research that exists on how benchmark assessments are used by teachers in the classroom, by principals, and by districts, as well as the
impact of benchmark assessments on student performance, there are even fewer largescale studies on the impact of benchmark assessments on student learning. In a Regional Educational Laboratory Northeast and Islands (REL-NEI) study, Henderson and colleagues explored the effect of quarterly benchmark exams in mathematics on 8th grade test scores using an interrupted 2-year time-series design, which produced non-significant gains in comparison to schools that did not implement benchmark assessment programs (Henderson et al., 2008). In a comparable study from Manpower Demonstration Research Corporation (MDRC), researchers studied the impact of benchmark assessments on student thinking in reading (FAST-R) on 3rd- and 4th grade state test scores and on SAT-9 performance using a similar interrupted time-series design, which produced generally positive but not statistically significant results (Chrismer, 2005).

**Concerns regarding benchmark assessment data.** The effectiveness of benchmark assessments, implementation, and data result usage is determined by teacher, principal, and district knowledge. Benchmark data are not being used by teachers to understand the extent to which students really understand content (Nabors Olah, 2010). Additionally, both teachers and school administrators have limited knowledge in evaluating the quality of assessment data (Bursch et al, 2010) as well as lack the knowledge, resources, and support to connect assessment results to teaching and learning (Datnow et al., 2007; Kerr et al., 2006; & Young, 2006). Moreover, teacher acceptance of data usage is low for those who have not been trained, prepared, or supported well in
using assessment data to modify instruction as they often become frustrated and resistance with analyzing student data (Trimble et al., 2005).

Blanc, Christman, Liu, Mitchell, Travers, and Bulkley (2010) emphasized “…student learning will not improve unless attention is paid to developing strong school leaders who promote data-driven decision making within a school culture focused on strengthening instruction, professional learning, and collective responsibility for student success” and that “…a well-designed system of interim assessments is dependent on the knowledge and skills of the school leaders and teachers who are responsible for bringing the system to life in schools” (Blanc et al., 2010, p. 206).

**Why Benchmark Testing?**

School districts throughout the nation concerned with how students will perform on the end-of-the year state tests are increasingly administering benchmark assessments (Olson, 2005; Sawchuk, 2009). Many recognize that information from annual state tests come too late in the school year to make a difference in student achievement; thus, leaving some students behind (Henderson, 2008; Herman, 2005; Herman et al., 2010; Pane et al., 2007). Approximately 70% of school superintendents in the nation reported their districts are using benchmark assessments and 10% plan to do so in the near future (Olson, 2005). Eighty percent of superintendents in California, Georgia, and Pennsylvania support results from benchmark assessments to be more useful for decision making than state test results (Marsh et al., 2007).
Benchmark assessments serve as a bridge to high stakes end-of-the-year state testing, minimizing the negative effects of feelings of powerlessness for teachers and hopelessness for students (Gorlewski, 2008). It is used throughout the year to measure and monitor student progress and provide teachers with data to transform instruction prior to students taking the state tests (Regional Educational Laboratory, 2007; Olson, 2005). Benchmark assessments give teachers immediate feedback on how and if their students are learning (Fournier et al., 2009). Some benchmark assessments provide extensive reporting systems that break down test results by the same student categories required under the federal No Child Left Behind Act (NCLB) of 2001 which includes race, income, disability, and English proficiency, along with providing individual district, school, classroom, and student progress reports (Olson, 2005).

**Research on impact of benchmark testing.** According to WestEd Senior Research Associate Susan Henderson (2008), not a lot of research exists on the impact of benchmark assessments. Herman and Baker (2005) further support that information about the effectiveness of benchmark assessments in improving student learning is generally hard to come by. Existing research supports the importance of benchmark assessments in classrooms due to its positive possibilities with respect to both policy and pedagogy (Gorlewski, 2008). In recent years, meta-analyses of research studies show that benchmark assessments have the potential to improve student learning (Black & Wiliam, 1998; Black et al., 2003; Kluger & DeNisi, 1996). As determined in Bancroft’s (2009) multi-year ethnographic study of English classrooms at an urban charter high
school serving mostly low-income minority students, low student skills and few resources (i.e. time, small group or one-on-one instruction) can undermine the efficacy and functionality of an effective benchmark assessment.

**Positive impact of benchmark testing.** Assessments can be used to forecast future performance (Perie et al., 2007). Using benchmark assessments to track student achievement and transform instructional practices to meet student needs have been found to be one important feature of high achieving districts according to studies by the Washington-based Council of the Great City Schools, the Austin, Texas-based National Center for Educational Accountability and others (Olson, 2005). In Forsyth County, administrators and teachers credit benchmark assessments combined with collegial data and best practice conversations for making impressive growth and having 16 elementary schools and eight middle schools making AYP (Pijanowski, 2008). In 2010, Blanc et al. conducted a multi-method study of benchmark assessments which examined how elementary school leaders and grade level teams in the Philadelphia School District make sense of benchmark data. The researchers concluded there was substantial evidence that benchmark assessments have the potential to contribute to instructional coherence and improvement that strengthened student success (Blanc et al., 2010).

**Negative impact of benchmark testing.** The effectiveness of benchmark testing in improving student learning is limited (Herman & Baker, 2005). Accordingly, teachers in Bancroft’s (2009) three-year investigation of low-income high school administration’s use of standardized benchmark assessments in English language arts negatively viewed
benchmark testing. They resented time that benchmark testing consumed from classroom instruction and its inability to adequately measure student progress (Bancroft, 2009). Furthermore, Herman and Baker (2005) find little sense in spending time and money on an elaborate testing system if the tests do not produce accurate and useful information needed to increase student achievement. Moreover, if benchmark tests do not reflect state standards and assessments, their results do not yield information about whether students are making adequate progress towards achieving the standards (Herman & Baker, 2005).

**Curriculum Associates**

Curriculum Associates [CA] (2011) publishers have published research-based, effective classroom-proven materials developed to help educators address the diverse learning needs of all students for the past 40 years. They have stated their materials are “Proven solutions for student success” and makes “…a profound difference in the lives of students and educators” as their “…programs are designed to meet state-specific standards to help schools across the country meet the rigorous demands of AYP” (http://www.curriculumassociates.com/products/detail.asp?title=CACSela, 2011).

**Product overview.** One of CA benchmark assessments product, Curriculum Associates California Content Standards and Mastery Series (CCS), is the focus of this study. It is aligned with California Standardized Test (CST) standards and formats. Each CCS content-area booklet contains four practice tests with items specifically correlated to the CST blueprints. Practice tests provide teachers the necessary diagnosis needed to
plan instructional strategies to target student learning deficiencies, which will ultimately lead to increased CST student proficiency levels.

**Research on CCS.** According to a study on CCS, results from CCS positively correlated with results on CST, translating to CCS being a good performance predictor for CST success (WEBSITE, YR.). In this study, 80% of students who scored 58% or higher on the CCS and received focused instruction scored proficient or advanced on the CST when they took the CCS 6-8 weeks prior to the CST testing window. Additionally, students who scored 45-57% on the CCS and received focused instruction had a highly probability of reaching proficiency.

ERIA (2011) explored whether CCS in English-language arts provided valid and reliable information teachers could use to plan focused and effective instruction. ERIA used data provided by 16 schools in a Northern California school district that had approximately 7300 students, grades 1 to 8 from 2009-2010 academic year. Approximately 21,900 tests (three different tests administered per student) were scored and analyzed for this study. Fifty-five percent of the students received free or reduced lunch, 62% Hispanic students, and 9% other minority students.

The results from this study concluded both validity and reliability (.86-.90) of CCS in English-language arts for each grade level were extremely high. First, all three tests taken at each grade level (Practice 1-3) were extremely reliable, implying teachers could use the results for instructional decision making. Second, student scores demonstrated consistent and statistically significant growth over the course of an
academic year. Third, there was evidence of a strong correlation between scores received on the CCS and scores received on the CST. Last, CCS results provided strong and consistent predictions from Practice 3 to the CST.

Students take CCS benchmark assessments throughout the year to demonstrate their mastery of grade-level standards. According to Ortland (2005), the benchmark is a comprehensive formative assessment that incorporates nearly all of a particular grade level standard. The benchmark is focused only on standards previously taught and given at strategic points in time. Ortland, further remarks that benchmark results provide teachers with critical information necessary in making key instructional decisions that can increase student learning and achievement.

Results from CCS benchmark assessments inform teachers about how and to what degree their instructional practices influence their students’ achievement as well as identify which essential standards teachers must focus their instruction to increase student learning and mastery (Ortland, 2005). Furthermore, the results provide teachers opportunities to reflect on the effectiveness of their instructional practices and make modifications to their instructional lens to better meet the learning needs of their students (Ortland, 2005). CCS data allows teachers to know if and to what extent students are making progress toward particular standards and may be used to adjust instruction to increase student success (Ortland, 2005).

CCS benchmark assessments have multiple assessments in each student booklet to allow students multiple opportunities to demonstrate their level of mastery on their grade
level standards (Ortland, 2005). Ortland indicated it is imperative that students take CCS benchmark assessments two to three times during the year in order for them to exhibit their learning growth over time.

CCS benchmark assessments are extremely thorough with rigorous deconstructed state standards closely aligned to the state standards test with the same number of items for a single standard as in the blueprint (Ortland, 2005). Students must understand and accurately process the academic language in the question stem to achieve a high level of success (Ortland, 2005). When compared with Bloom’s Taxonomy, CCS benchmark questions are beyond the knowledge level and reflective of high-order thinking skills that require students to apply their comprehension.

Ortland identified the essence of each piece of CCS benchmark assessment data. Data from the first benchmark assessment provides teachers with a baseline of incoming student performance and achievement in that grade level standards. It allows teachers to focus on prevention versus intervention for students who come in from the previous grade at below proficiency level (Ortland, 2005). She specified how teachers may strategically focus their instruction on particular essential standards and for whom. The second benchmark may be given after approximately sixty days of instruction. Data from this benchmark provide additional information for comparative growth (Ortland, 2005). Ortland determined teachers may strategically focus their instruction on student intervention needs as well as provide additional support on specific standards. Data from the last benchmark provide teachers with a focus on intervention. Ortland emphasized
the importance of this piece of data in critically identifying non-proficient students and engaging them in individualized instruction focused on standard mastery.

**Conclusion**

Though there is extensive discussion in the literature on the impact of high stakes testing, specific research on benchmark testing is limited. More specifically, there is a lack of research base concerning its effectiveness. This study will add to the literature through an examination of the relationship between benchmark testing and statewide assessments. In particular, the study will focus on the influence of benchmarking testing on teaching practices of teachers and instructional practices of principals. The results of this study will be used to discuss the implications for instruction and educational leadership at Title I Program Improvement elementary schools.
Chapter 3
METHODOLOGY

Despite the lack of research base concerning its effectiveness (Henderson et al., 2007), many school districts and schools nationwide are administering periodic benchmark assessments to complement the end-of-year state testing and increase student learning as a response to accountability pressures to raise student achievement. The question remains whether there is a significant difference between scores on benchmark assessments and scores on state test? Likewise, how do benchmark assessments impact teaching practices of teachers and leadership practices of principals at Title I Program Improvement (PI) schools?

This study examined the relationship between benchmark assessments and the end-of-year state testing in two PI Title I schools in one school district in northern California (NCSD). More specifically, this study investigated whether there is a significant difference between student scores on the final benchmark assessment and student scores on state tests taken at a later date (Mid-Atlantic Regional Educational Laboratory, 2007) and attempted to answer the following research questions:

1. Is there a significant difference between students’ Curriculum Associates Benchmark (CAB) scores and students’ California Standards Test (CST) scores?
a. Is there a significant difference between student scores on the final benchmark assessment in English Language Arts and student outcomes on the California Standards Test?

b. Is there a significant difference between student scores on the final benchmark assessment in mathematics and student outcomes on the California Standards Test?

c. Are student scores on the final benchmark assessments in English Language Arts and student outcomes on the California Standards Test significantly different for student subgroups (EL, Non-EL, African American, Asian, and Latino)?

d. Are student scores on the final benchmark assessments in mathematics and student outcomes on the California Standards Test significantly different for student subgroups (EL, Non-EL, African American, Asian, and Latino)?

e. Are student scores on the final benchmark assessments in English Language Arts and student outcomes on the California Standards Test significantly different by grade levels (grades 2, 3, 4, and 5)?

f. Are student scores on the final benchmark assessments in mathematics and student outcomes on the California Standards Test significantly different by grade levels (grades 2, 3, 4, and 5)?
2. In what ways have benchmark assessments influenced teaching practices of teachers in elementary Year 5 Title I Program Improvement schools?
   a. How have benchmark assessments influenced teachers to modify their instructional approaches?
   b. How have benchmark assessments impacted the ways in which teachers plan their lessons?
   c. How have benchmark assessments changed the ways in which teachers work with specific groups of students?
   d. What influence have benchmark assessments had on teacher practices regarding evaluation of student achievement?

3. In what ways have benchmark assessments influenced the instructional leadership practices of principals in elementary Year 5 Title I Program Improvement schools?
   a. How have benchmark assessments influenced principals to modify their leadership approaches?
   b. How have benchmark assessments impacted the ways in which principals monitor instructions?
   c. How have benchmark assessments changed the ways in which principals work with teachers?

This chapter describes the research methodology used in the study. In particular, the setting of the study, population, sampling methods, materials used, data collection
approach, data analysis, and discussion of the protection of participants’ rights are discussed.

Research Design and Approach

This study utilized a mixed methods research design, which helped to connect quantitative and qualitative research (Onwuegbuzie & Leech, 2004a) as both are important and valuable measures (Johnson & Onwuegbuzie, 2004). Johnson and Onwuegbuzie defined mixed methods research as “the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, and concepts or language in a single study” (2004, p. 7). Mixed methods was chosen for this study because of its goal “…to draw from the strengths and minimize the weaknesses of both types of research in single research studies and across studies” (Johnson & Onwuegbuzie, 2004, p. 14-15). More specifically, benchmark data and CST data were analyzed, and teacher and principal surveys and interviews were used.

Quantitative Study

The quantitative portion of this study was used to answer research questions one through three. The first research question focused on the relationship between the independent variable (CAB score) and the dependent or outcome variable (CST score). This quantitative research design is descriptive and establishes the relationship between the two aforementioned variables. In order to make an accurate assumption about the relationship between the independent and dependent variables, this descriptive study used
a sample of 317 students (for questions 1a and 1b) as the estimate of the relationship is less likely to be biased as the number of randomly selected participants is high (Hopkins, 2000). The second and third research questions focused on the influence of benchmark testing on teaching practices of teachers and instructional leadership practices of principals. The participants were asked to complete a four-point Likert-like survey attesting to their level of practices.

**Qualitative Study**

The qualitative portion of this study was used to answer research questions two and three. The aim of this research design emphasized the importance of looking at variables in a natural setting. Interaction between variables is important. Detailed data is gathered through open-ended questions closely linked to the survey questions along with use of direct quotations. The qualitative portion of this research contained interview narrative data that sought details, clarification, and depths into teaching or instructional leadership practices pertaining to benchmark testing. Qualitative research was used in this study to gain insight into the participants’ attitudes, behaviors, concerns, motivations, and value systems with reference to benchmark assessments and state tests. This descriptive research will be used to inform decisions within schools (particularly Title I PI schools), school and district policy, and research. This method was used to seek understanding of people’s interpretations of their practices.
Setting and Sample

Northern California School District (NCSD) is a third year PI urban school district. The school district is one of the 10 largest school districts in the western states. Within NCSD, there are six Title I PI Year 5 public schools. Title I schools are low-achieving schools that have a high percentage of students from low income families. Title I is the oldest and largest federally funded program in the nation in support of students at risk of failure or living at or near poverty. This policy, committed to closing the achievement gap between low income students and other students, was originally enacted in 1965 under the Elementary and Secondary Education Act (ESEA) but was rewritten in 1994 to improve essential goals of helping at-risk students.

The six schools in NCSD together are comprised of approximately 2,632 ethnically and linguistically diverse students from low income households. The three largest ethnic groups in these six elementary schools are African Americans, Asians, and Latinos. The schools have API growth scores ranging from 656-781. Figure one offers detailed information for each of the schools. Each school is expected to have a range of 270-564 students. School A has 424 students; School B has 387 students; School C has 270 students; School D has 440 students; School E has 429 students; and School F has 564 students.
Table 1: School Demographics

<table>
<thead>
<tr>
<th>School</th>
<th>API</th>
<th>Enrollment</th>
<th>African Americans</th>
<th>Asians</th>
<th>Latino Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>781</td>
<td>424</td>
<td>11%</td>
<td>12%</td>
<td>67%</td>
</tr>
<tr>
<td>B</td>
<td>656</td>
<td>387</td>
<td>26%</td>
<td>37%</td>
<td>28%</td>
</tr>
<tr>
<td>C</td>
<td>665</td>
<td>270</td>
<td>53%</td>
<td>19%</td>
<td>19%</td>
</tr>
<tr>
<td>D</td>
<td>692</td>
<td>440</td>
<td>28%</td>
<td>40%</td>
<td>25%</td>
</tr>
<tr>
<td>E</td>
<td>736</td>
<td>429</td>
<td>13%</td>
<td>7%</td>
<td>61%</td>
</tr>
<tr>
<td>F</td>
<td>744</td>
<td>564</td>
<td>8%</td>
<td>37%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Quantitative Study

There are six schools within the district that are categorized as Title I, PI Year 5. The researcher requested to study all six schools in the district. However, only two of the six principals agreed to share their test data and allow their teachers to participate in the study. As a result, the NCSD provided the researcher with student data from only two schools. The CAB data consisted of scores for 356 students and the CST data consisted of scores for 588 students. The researcher wanted to include students who had both CAB scores and CST scores. She compared all students who had a CST score to students who had a CAB score. Students who had a CST score, but did not have a CAB score, were not included in this study. The result of this process provided the researcher with 317 students to use in this study for research question 1a and 1b (Refer to Table 2). For
research questions 1c-1d, the researcher deleted all students that did not fall into each specific subgroup used in this study, which resulted in 77 African American, 101 Asian, 99 Latino, 150 English Learners (EL), and 166 Non-EL students (Refer to Tables 4 and 5). Likewise, for research questions 1e-1f, the researcher concluded there were 77 second grade, 82 third grade, 80 fourth grade, and 77 fifth grade students (Refer to Table 3).

*Table 2: Number of Students Studied*

<table>
<thead>
<tr>
<th>School</th>
<th>N</th>
<th>Percentage of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>124</td>
<td>39%</td>
</tr>
<tr>
<td>School B</td>
<td>193</td>
<td>61%</td>
</tr>
<tr>
<td>Total</td>
<td>317</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Table 3: Number of Students Studied by Grade*

<table>
<thead>
<tr>
<th>School</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>35</td>
<td>28</td>
<td>35</td>
<td>30</td>
<td>128</td>
</tr>
<tr>
<td>School B</td>
<td>43</td>
<td>54</td>
<td>45</td>
<td>47</td>
<td>189</td>
</tr>
<tr>
<td>Total</td>
<td>78</td>
<td>82</td>
<td>80</td>
<td>77</td>
<td>317</td>
</tr>
</tbody>
</table>
Table 4: Number of Students Studied by Ethnicity

<table>
<thead>
<tr>
<th>School</th>
<th>African American</th>
<th>Asian</th>
<th>Latino</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>32</td>
<td>38</td>
<td>36</td>
<td>106</td>
</tr>
<tr>
<td>School B</td>
<td>46</td>
<td>65</td>
<td>63</td>
<td>174</td>
</tr>
<tr>
<td>Total</td>
<td>78</td>
<td>103</td>
<td>99</td>
<td>280</td>
</tr>
</tbody>
</table>

Table 5: Number of Students Studied by Language Proficiency

<table>
<thead>
<tr>
<th>School</th>
<th>EO</th>
<th>EL</th>
<th>Percentage of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>70</td>
<td>54</td>
<td>124</td>
</tr>
<tr>
<td>School B</td>
<td>96</td>
<td>97</td>
<td>193</td>
</tr>
<tr>
<td>Total</td>
<td>165</td>
<td>151</td>
<td>317</td>
</tr>
</tbody>
</table>

Qualitative Study

The qualitative portion of this study was used to answer research questions two and three. The aim of this research design emphasized the importance of looking at variables in a natural setting as interaction between variables is important. Detailed data is gathered through nine open-ended questions closely linked to the survey questions along with use of direct quotations. The qualitative portion of this research contained interview narrative data that sought to answer why by analyzing unstructured information along with two open-ended questions located at the end of the teacher and principal survey. Qualitative research was used to gain insight into the participants’ attitudes,
behaviors, concerns, motivations, and value systems with reference to benchmark assessments and the CST. This research will be used to inform decisions within schools (particularly Title I PI schools), school and district policy, and research. The qualitative portion of this study used a small group of nine teachers and one principal to guide and construct hypotheses. The results are descriptive. Furthermore, this portion of the research included in-depth interviews (9 questions) with the nine teachers and no principal. The 10-15 minute sessions were conducted in person or via telephone, depending on the participant’s choice. One interview was conducted with a group of two teachers, while five interviews were conducted individually in the researcher’s office, and two interviews were conducted via telephone from the researcher’s home. With the exception of the interviews being done via telephone, all interview participants were provided a copy of the interview questions to access during the interview. The purpose of this method was to seek understanding of people’s interpretations.

**Instrumentation and Materials**

Several tools were used to answer the research questions in this study: benchmark assessment and state test data, survey questionnaire, open-ended questions on the survey, and interview. Student test scores for English language arts (ELA) and math for CAB and CST were used for the quantitative portion of the study to answer question one (1a-1f). The researcher converted the proficiency level text earned by students on the CAB to
a numerical score to align with established CST scores, which resulted in the following: Advanced (5), Proficient (4), Basic (3), Below Basic (2), and Far Below Basic (1).

In addition to analyzing if there was a significant difference between CAB scores and CST scores, 17 out of 23 teachers and one out of two principals from two Title I PI Year 5 schools were surveyed using a structured questionnaire that contained predominantly 14 closed-ended or forced choice responses, which were used to answer the quantitative portion of questions two and three. A forced answer closed-choice four-point Likert-style survey was also used for the quantitative portion of the study to answer questions two. The teacher survey (see Appendix A) collected background information pertaining to where the teacher teaches, what grade they teach, how long they have been teaching in their current grade level, and how long they have been teaching altogether. The forced choice part of the survey asked teachers to rate to what extent they agreed to each statement.

The following are examples of the teacher survey: “I have made changes to my instructional approach as a result of benchmark assessments;” “I use benchmark assessments to create activities for my students to do as an extension of my instruction;” “My instructional approach is more focused on standards covered by benchmark assessments,” “I use benchmark results to plan and organize differentiated lessons to meet my students’ needs;” and “I spend more time working with specific groups of students than I have before because of benchmark assessments.”
In addition to the teacher survey, a principal survey (see Appendix B) was also used for the quantitative portion of the study to answer question three. The principal survey collected background information pertaining to how long they have been principal at their current school, how many years they have served as principal, and their experience with implementing benchmark assessments as a school administrator as well as a classroom teacher. The forced choice part of the survey asked principals to rate to what extent they agreed to each statement. The following are examples of the principal survey: “My classroom observations are focused on the integration of benchmark assessments into the curriculum;” “My classroom observations are focused on the use of benchmark results to meet students’ specific needs;” “My classroom observations are focused on standards covered in benchmark assessments;” “I hold meetings and trainings on how teachers can use benchmark assessments in their lesson planning;” and “My instructional discussion with teachers is more focused on standards covered in benchmark assessments.”

As for the qualitative portion of the study, two open-ended questions at the end of the teacher and principal survey were presented. The questions at the end of the survey asked, “What are some additional positive features of the benchmark assessments that have been useful for you?” and “What are some shortcomings/challenges as to the use of benchmark assessments?”

Along with the two open-ended questions at the end of the survey, teacher interviews were also conducted to support the qualitative portion of the study. The
following are examples of questions participants were asked: “How have you changed your instructional approach as a result of benchmark assessments;” “What kinds of activities have you created as a result of benchmark assessments for your students to do as an extension of your instruction;” “How is your instructional approach more focused on standards covered by benchmark assessments;” “How do you use benchmark assessments to plan and organize small group instruction;” and “How have benchmark results impacted your approach to teaching specific groups of students?”

**Data Collection Procedures**

The quantitative portion of the study examined the relationship between an independent variable (CAB score) and a dependent or outcome variable (CST score). The data was provided by a research specialist in the Assessment, Research, and Evaluation (ARE) Department of NCSD. The researcher disaggregated the data to exclude special education students as they take a modified version of the CST. Also, the researcher disaggregated the data to answer questions two and three for African American, Asian, and Latino student subgroups as well as students by grade levels to determine if there was a significant difference for student scores on the CAB and CST for each group. The researcher assigned an alphabetical code for each of the selected schools in the study to reserve their anonymity.

In addition to examining student scores on the CAB and CST, the researcher also had the 17 teachers and one principal in the study complete a four-point Likert-style
survey to support the quantitative portion of the study. The researcher administered the survey during a staff meeting and collected the survey prior to leaving the meeting. Each survey was then assigned a number and compiled into an Excel document.

The qualitative portion of the study covered how benchmark assessments influenced teaching practices of teachers and instructional leadership practices of principals. The researcher interviewed teachers at both schools to further investigate how benchmark assessments impacted their teaching practices (See Appendix C), but no interview was completed with either principal as neither gave their consent. Two of the teacher interviews were conducted via telephone and seven were completed in person. The researcher provided the participants with a copy of the interview questions and hand wrote their responses on another copy of the interview questions. The interview took between 10-15 minutes to complete.

Data Analysis

SPSS (Statistical Package for the Social Sciences), a data analysis program designed for social science research, was utilized to calculate several paired-samples t test to determine if there was a significant difference between two variables (student scores on the benchmark and state test). Initially, research question one sought to determine the “correlational” relationship between students’ CAB and CST scores. However, upon inputting data, the researcher observed the scores were positively correlated and felt it was more appropriate to investigate the significant difference
between the scores, resulting in a paired-samples \( t \) test versus the initially planned independent-samples \( t \) test. The paired-samples \( t \) test evaluate within group significance (Green & Salkind, 2008). The within group analyses involve evaluating significance of student scores on the CAB and CST for all students who have both a CAB score and a CST score, subgroups (African American, Asian, Latino, EL, and Non-EL), and by grade level (grades 2, 3, 4, and 5). The paired-samples \( t \)-test was used to answer research question one (1a-1f).

The qualitative portion of this study analyzed teacher interviews regarding the role of benchmark assessments on their teaching practices as well as the two open-ended questions at the end of the survey. The researcher used a qualitative research design to gain a better understanding of what ways (if any) benchmark assessments impact teaching practices. Qualitative research helps in understanding and explaining social phenomena (Merriam, 1998). The researcher expected, by using a qualitative research design, to identify specific aspects of benchmark assessments that contributed to student success on the CST.

**Protection of Participants**

The subject of this research, the relationship between benchmark assessments and CST tests, is not a particularly sensitive subject, so low risk to participants was foreseen. Despite the anticipated low risk, the researcher protected the participants and maintained the confidentiality of their responses to the surveys and interviews.
No student name was used in the student data provided to the researcher by the district’s research specialist. Each student was already assigned a number identifier for both sets of scores (CAB and CST). Additionally, the researcher assigned each survey and interview (respondents or interviewee) a number identifier. The surveys and interviews were kept in a locked cabinet in the researcher’s home and no identifying information was used in the written results and student test scores are stored in the researcher’s computer desktop, which requires a password to access. This research project was approved by the California State University, Sacramento Committee for the Protection of Human Subjects and the school district (NCSD).

**Conclusion**

This study used student scores from the last benchmark and end-of-year state test to examine whether there is a significant difference between student scores on benchmark assessments (CAB) and state tests (CST) along with teacher and principal survey and interview to determine the role of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals. Data from two of the six PI schools in the district were used in this study. Analysis of this data has the potential to provide teachers and educational leaders with clear insights into benchmark testing, its role in teaching, learning, and leadership practices, as well as potential impact on increasing student achievement on high stakes state standardized tests.
Chapter 4
ANALYSIS OF THE DATA

This chapter reported findings for the study and was structured around the research questions addressed. It provided a data analysis of the research questions and underlying theoretical framework of the study. The chapter concluded with a brief summary of the outcomes and interpreted relationship to the research questions.

The first section provided an analysis of the quantitative data collected in the study—student scores on benchmark assessments (CAB) and annual state tests (CST) along with teacher and principal surveys. This section presented findings on whether there is a significant difference between student scores on the Curriculum Associates Benchmarks (CAB) and student scores on the California Standards Tests (CST). It also reported on teacher and principal perception on the role of benchmark assessments. The proceeding section concluded with an analysis of the qualitative data collected in the study—teacher interviews on the role of benchmark assessments on their teaching practices. The findings in this study were intended to answer the following research questions:

1. Is there a significant difference between Curriculum Associates Benchmark (CAB) scores and English language arts California Standards Test (CST) scores?
a. Is there a significant difference between student scores on the final benchmark assessment in English Language Arts and student outcomes on the California Standards Test?
b. Is there a significant difference between student scores on the final benchmark assessment in mathematics and student outcomes on the California Standards Test?
c. Are student scores on the final benchmark assessments in English Language Arts and student outcomes on the California Standards Test statistically significant for student subgroups?
d. Are student scores on the final benchmark assessments in mathematics and student outcomes on the California Standards Test statistically significant for student subgroups?
e. Are student scores on the final benchmark assessments in English Language Arts and student outcomes on the California Standards Test significantly different by grade levels?
f. Are student scores on the final benchmark assessments in mathematics and student outcomes on the California Standards Test statistically significant by grade levels?

2. In what ways have benchmark assessments influenced teaching practices of teachers in elementary Year 5 Title I Program Improvement schools?
a. How have benchmark assessments influenced teachers to modify their instructional approaches?

b. How have benchmark assessments impacted the ways in which teachers plan their lessons?

c. How have benchmark assessments changed the ways in which teachers work with specific groups of students?

d. What influence have benchmark assessments had on teacher practices regarding evaluation of student achievement?

3. In what ways have benchmark assessments influenced the instructional leadership practices of principals in elementary Year 5 Title I Program Improvement schools?

a. How have benchmark assessments influenced principals to modify their leadership approaches?

b. How have benchmark assessments impacted the ways in which principals monitor instructions?

c. How have benchmark assessments changed the ways in which principals work with teachers?
Data Analysis

Quantitative Study: CAB and CST Scores

This research focused on two of six Title I PI Year 5 schools within one of the ten largest school districts in the western states. For this study, paired-samples $t$ tests were used to compare the means for student scores on the CAB and student scores on the CST. The researcher converted the proficiency level text earned by students on the CAB to a numerical score to align with established CST scores, which resulted in the following: Advanced (5), Proficient (4), Basic (3), Below Basic (2), and Far Below Basic (1).

For the purpose of this study, student scores for the school year 2010-2011 were reviewed. The findings of these analyses are presented in this section. First, the results of all students (excluding special education students and students missing either test score), are reported. Then, results of the analyses related to particular subgroups such as African American, Asian, Latino, EO and non-EO students, along with each grade level (grades 2, 3, 4, and 5) are presented.

All students. The first analysis determined whether the mean value of the test scores for all students who took the CAB differed significantly from the mean value of the test scores of all students who took the CST. The number of students for this portion of the study totaled 317. Table 6 illustrates the results of this analysis.
Table 6: Mean Value of Test Scores for All Students

<table>
<thead>
<tr>
<th>Test Instrument</th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>2.87</td>
<td>3.00</td>
</tr>
<tr>
<td>CST</td>
<td>2.90</td>
<td>3.24</td>
</tr>
<tr>
<td>Difference between CAB and CST</td>
<td>-.03</td>
<td>-.24</td>
</tr>
</tbody>
</table>

Table 6 shows all students who took the CAB and CST had minimal increase in their CST scores in ELA (from 2.87 to 2.90, resulting in a difference of -.03) but a significant difference in the math portion of the test (from 3.00 to 3.24, resulting in a difference of -.24). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for the ELA portion of the test, but a significant difference for the math portion of the test. The results indicated that the mean CABela and CSTela ($M = -.02215, SD = .88164$) was significantly greater than the mean for CABmath and CSTmath ($M = -.24684, SD = .99002$), $t(315) = 4.43, p < .001$). The standardized effect size index, $d$, was .03 for CABela and CSTela and .25 for CABmath and CSTmath, with considerable overlap in the distributions for the 5-point Likert ratings of CAB and CST. The 95% confidence interval for the mean difference between the CABela and CSTela was -.12 and +.08 and CABmath and CSTmath was -.36 and -.14.

When focusing exclusively on the CAB score and CST score for all students, there are some notable results in the math test. To see if there were significant differences in student scores on the CAB and the CST, the researcher used a paired
samples $t$-test to evaluate within group comparisons. Table 7 shows the results of this analysis.

*Table 7: Paired Samples Test for All Students*

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAB ELA &amp; CST ELA</td>
<td>CAB Math &amp; CST Math</td>
</tr>
<tr>
<td>Mean</td>
<td>-.02215</td>
<td>-.24684</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.88164</td>
<td>.99002</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td>.04960</td>
<td>.05569</td>
</tr>
<tr>
<td>95% Confidence Interval of the Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>-.11973</td>
<td>-.35641</td>
</tr>
<tr>
<td>Upper</td>
<td>.07543</td>
<td>-.13726</td>
</tr>
<tr>
<td>t</td>
<td>-.447</td>
<td>-4.432</td>
</tr>
<tr>
<td>df</td>
<td>315</td>
<td>315</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.655</td>
<td>.000</td>
</tr>
</tbody>
</table>

**African American students.** The next analysis determined whether the mean value of the test scores for all African American students who took the CAB differed significantly from the mean value of the test scores of all African American students who took the CST. The number of African American students for this portion of the study totaled 76. Table 8 establishes the results of this analysis.

Table 8 shows all African American students who took the CAB and CST had fairly minimal increase in their CST scores in ELA (from 2.71 to 2.92, resulting in a
difference of -.21) but a rather significant difference in the math portion of the test (from 2.63 to 3.17, resulting in a difference of -.54). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for the ELA portion of the test, but a significant difference for the math portion of the test. The results indicated that the mean CABela and CSTela \((M = -0.20779, SD = .80029)\) was significantly greater than the mean for CABmath and CSTmath \((M = -0.53247, SD = 1.22027), t(76) = 3.829, p < .001)\). The standardized effect size index, \(d\), was .26 for CABela and CSTela and .44 for CABmath and CSTmath with considerable overlap in the distributions for the 5-point Likert ratings of CAB and CST. The 95% confidence interval for the mean difference between the CABela and CSTela was -39 and -0.03 and CABmath and CSTmath was +.81 and -0.26.

*Table 8: Mean Value of African American Students*

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>2.71</td>
<td>2.63</td>
</tr>
<tr>
<td>CST</td>
<td>2.92</td>
<td>3.17</td>
</tr>
<tr>
<td>Difference</td>
<td>-.21</td>
<td>-.54</td>
</tr>
</tbody>
</table>

When focusing exclusively on the CAB score and CST score for all African American students, there are some notable results in the math test. To see if there were significant differences in African American student scores on the CAB and on the CST,
the researcher used a paired samples *t*-test to evaluate within group comparisons. Table 9 displays the results of this analysis.

*Table 9: Paired Samples Test for African American Students*

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAB ELA &amp; CST ELA</td>
<td>CAB Math &amp; CST Math</td>
</tr>
<tr>
<td>Mean</td>
<td>-.20779</td>
<td>-.53247</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.80029</td>
<td>1.22027</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td>.09120</td>
<td>.13906</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>-.38944</td>
<td>-.80944</td>
</tr>
<tr>
<td>Lower</td>
<td>-.02615</td>
<td>-.25550</td>
</tr>
<tr>
<td>Upper</td>
<td>-2.278</td>
<td>-3.829</td>
</tr>
<tr>
<td>t</td>
<td>-2.278</td>
<td>-3.829</td>
</tr>
<tr>
<td>df</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.026</td>
<td>.000</td>
</tr>
</tbody>
</table>

Asian students. The following analysis determined whether the mean value of the test scores for all Asian students who took the CAB differed significantly from the mean value of the test scores of all Asian students who took the CST. The number of Asian students for this portion of the study totaled 100. Table 10 establishes the results of this analysis.
Table 10: Mean Value of Asian Students

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>3.08</td>
<td>3.29</td>
</tr>
<tr>
<td>CST</td>
<td>2.94</td>
<td>3.57</td>
</tr>
<tr>
<td>Difference</td>
<td>+.14</td>
<td>-.28</td>
</tr>
</tbody>
</table>

Table 10 shows all Asian students who took the CAB and CST had minimal decrease in their CST scores in ELA (from 3.08 to 2.94, resulting in a difference of .14) but a significant difference in the math portion of the test (from 3.29 to 3.57, resulting in a difference of -.28). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for the ELA portion of the test, but a significant difference for the math portion of the test. The results indicated that the mean CABela and CSTela ($M = .14851$, $SD = .86471$) was significantly greater than the mean for CABmath and CSTmath ($M = -.28713$, $SD = .77893$), $t(100) = 3.705$, $p < .001$. The standardized effect size index, $d$, was .17 for CABela and CSTela and .37 for CABmath and CSTmath, with considerable overlap in the distributions for the 5-point Likert ratings of CAB and CST. The 95% confidence interval for the mean difference between the CABela and CSTela was -.02 and +.32 and CABmath and CSTmath was -.44 and -.13.

When focusing exclusively on the CAB score and CST score for all Asian students, there are some notable results in the math test. To see if there were significant differences in Asian student scores on the CAB and the CST, the researcher used a paired
samples $t$-test to evaluate within group comparisons. Table 11 shows the results of this analysis.

*Table 11: Paired Samples Test for Asian Students*

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Standard Error of Mean</th>
<th>95% Confidence Interval Lower</th>
<th>95% Confidence Interval Upper</th>
<th>$t$</th>
<th>df</th>
<th>Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1: CAB ELA &amp; CST ELA</td>
<td>.14851</td>
<td>.86471</td>
<td>.08604</td>
<td>-.02219</td>
<td>.31922</td>
<td>1.726</td>
<td>100</td>
<td>.087</td>
</tr>
<tr>
<td>Pair 2: CAB Math &amp; CST Math</td>
<td>-.28713</td>
<td>.77893</td>
<td>.07751</td>
<td>-.44090</td>
<td>-.13336</td>
<td>-3.705</td>
<td>100</td>
<td>.000</td>
</tr>
</tbody>
</table>

**Latino students.** The proceeding analysis determined whether the mean value of the test scores for all Latino students who took the CAB differed significantly from the mean value of the test scores of all Latino students who took the CST. The number of Latino students for this portion of the study totaled 98. Table 12 exhibits the results of this analysis.

Table 12 shows all Latino students who took the CAB and CST had minimal increase in their CST scores in ELA (from 2.78 to 2.80, resulting in a difference of -.02)
and decrease in math (from 2.96 to 2.90, resulting in a difference of +.06). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for both ELA and math.

When focusing exclusively on the CAB score and CST score for all Latino students, there are no notable results for both the ELA and math tests. To see if there were significant differences in Latino student scores on the CAB and the CST, the researcher used a paired samples $t$-test to evaluate within group comparisons. Table 13 indicates the results of this analysis.

*Table 12: Mean Value of Latino Students*

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>2.78</td>
<td>2.96</td>
</tr>
<tr>
<td>CST</td>
<td>2.80</td>
<td>2.90</td>
</tr>
<tr>
<td>Difference</td>
<td>-.02</td>
<td>+.06</td>
</tr>
</tbody>
</table>

**Non-EL students.** The subsequent analysis determined whether the mean value of the test scores for all non-EL students who took the CAB differed significantly from the mean value of the test scores of all non-EL students who took the CST. The number of non-EL students for this portion of the study totaled 165. Table 14 demonstrates the results of this analysis.
Table 13: Paired Samples Test for Latino Students

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAB ELA &amp;</td>
<td>CAB Math &amp;</td>
</tr>
<tr>
<td></td>
<td>CST ELA</td>
<td>CST Math</td>
</tr>
<tr>
<td>Mean</td>
<td>-.02020</td>
<td>.06061</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.93656</td>
<td>.85498</td>
</tr>
<tr>
<td>Standard Error of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.09413</td>
<td>.08593</td>
</tr>
<tr>
<td>95% Confidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>-.20699</td>
<td>-.10992</td>
</tr>
<tr>
<td>Upper</td>
<td>.16659</td>
<td>.23113</td>
</tr>
<tr>
<td>t</td>
<td>-.215</td>
<td>.705</td>
</tr>
<tr>
<td>df</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.831</td>
<td>.482</td>
</tr>
</tbody>
</table>

Table 14: Mean Value of Non-EL Students

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>2.94</td>
<td>2.98</td>
</tr>
<tr>
<td>CST</td>
<td>3.04</td>
<td>3.31</td>
</tr>
<tr>
<td>Difference</td>
<td>-.1</td>
<td>-.33</td>
</tr>
</tbody>
</table>

Table 14 shows non-EL students who took the CAB and CST had minimal increase in their CST scores in ELA (from 2.94 to 3.04, resulting in a difference of -.1) but a significant difference in the math portion of the test (from 2.98 to 3.31, resulting in
a difference of -.33). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for the ELA portion of the test, but a significant difference for the math portion of the test. The results indicated that the mean CABela and CSTela ($M = -.10241, SD = .75982$) was significantly greater than the mean for CABmath and CSTmath ($M = -.33735, SD = 1.08742$), $t(165) = 3.997, p < .001$). The standardized effect size index, $d$, was .14 for CABela and CSTela and .31 for CABmath and CSTmath, with considerable overlap in the distributions for the 5-point Likert ratings of CAB and CST. The 95% confidence interval for the mean difference between the CABela and CSTela was -.22 and +.01 and CABmath and CSTmath was -.5 and -.17.

*Table 15: Paired Samples Test for Non-EL Students*

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB ELA &amp; CST ELA</td>
<td>-.10241</td>
<td>-.33735</td>
</tr>
<tr>
<td>CAB Math &amp; CST Math</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error of Mean</th>
<th>95% Confidence Interval</th>
<th>t</th>
<th>df</th>
<th>Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.05897</td>
<td>.08440</td>
<td></td>
<td></td>
<td>-1.737</td>
<td>165</td>
<td>.084</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.75982</td>
<td>1.08742</td>
<td></td>
<td></td>
<td>-.5</td>
<td>165</td>
<td>.000</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.17071</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>-.21885</td>
<td>-.50399</td>
<td></td>
<td></td>
<td>-.3997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>-.01403</td>
<td>-.17071</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.084</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
When focusing exclusively on the CAB score and CST score for all non-EL students, there are some notable results in the math test. To see if there were significant differences in non-EL student scores on the CAB and the CST, the researcher used a paired samples $t$-test to evaluate within group comparisons. Table 15 shows the results of this analysis.

**EL students.** The analysis below determined whether the mean value of the test scores for all EL students who took the CAB differed significantly from the mean value of the test scores of all EL students who took the CST. The number of EL students for this portion of the study totaled 149. Table 16 reflects the results of this analysis.

Table 16 shows EL students who took the CAB and CST had minimal decrease in their CST scores in ELA (from 2.79 to 2.73, resulting in a difference of +.06) and increase in math (from 3.01 to 3.15, resulting in a difference of -.14). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for both ELA and math.

When focusing exclusively on the CAB score and CST score for all EL students, there are little notable results in both ELA and math tests. To see if there were significant differences in EL student scores on the CAB and the CST, the researcher used a paired samples $t$-test to evaluate within group comparisons. Table 17 expresses the results of this analysis.
Table 16: Mean Value of EL Students

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>2.79</td>
<td>3.01</td>
</tr>
<tr>
<td>CST</td>
<td>2.73</td>
<td>3.15</td>
</tr>
<tr>
<td>Difference</td>
<td>+.06</td>
<td>-.14</td>
</tr>
</tbody>
</table>

Table 17: Paired Samples Test for EL Students

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB ELA &amp; CST ELA</td>
<td>Mean</td>
<td>.06667</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>.99439</td>
</tr>
<tr>
<td></td>
<td>Standard Error of Mean</td>
<td>.08119</td>
</tr>
<tr>
<td></td>
<td>95% Confidence Interval Lower</td>
<td>-.09377</td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>.22710</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>.821</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>Significance (2-tailed)</td>
<td>.413</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CAB Math &amp; CST Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
</tr>
<tr>
<td>95% Confidence Interval Lower</td>
</tr>
<tr>
<td>Upper</td>
</tr>
<tr>
<td>t</td>
</tr>
<tr>
<td>df</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
</tr>
</tbody>
</table>

Second grade students. The ensuing analysis determined whether the mean value of the test scores for second grade students who took the CAB differed significantly from the mean value of the test scores of second grade students who took the CST. The
number of second grade students for this portion of the study totaled 76. Table 18 demonstrates the results of this analysis.

*Table 18: Mean Value of Second Grade Students*

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>3.01</td>
<td>3.44</td>
</tr>
<tr>
<td>CST</td>
<td>3.03</td>
<td>3.25</td>
</tr>
<tr>
<td>Difference</td>
<td>-.02</td>
<td>+.19</td>
</tr>
</tbody>
</table>

Table 18 shows second grade students who took the CAB and CST had minimal increase in their CST scores in ELA (from 3.01 to 3.03, resulting in a difference of -.02) and a minimal decrease in the math portion of the test (from 3.44 to 3.25, resulting in a difference of +.19). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for the ELA portion of the test, but a significant difference for the math portion of the test. The results indicated that the mean CABela and CSTela ($M = -.01299, SD = 1.05747$) was significantly less than the mean for CABmath and CSTmath ($M = .29870, SD = .74454$), $t(76) = 3.520, p < .001$). The standardized effect size index, $d$, was .01 for CABela and CSTela and .4 for CABmath and CSTmath with considerable overlap in the distributions for the 5-point Likert ratings of CAB and CST. The 95% confidence interval for the mean difference between the CABela and CSTela was -.25 and +.23 and CABmath and CSTmath was +.13 and +.47.
When focusing exclusively on the CAB score and CST score for second grade students, there are some notable results in the math test. To see if there were significant differences in second grade student scores on the CAB and the CST, the researcher used a paired samples $t$-test to evaluate within group comparisons. Table 19 exhibits the results of this analysis.

Table 19: Paired Samples Test for Second Grade Students

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAB ELA &amp; CST ELA</td>
<td>CAB Math &amp; CST Math</td>
</tr>
<tr>
<td>Mean</td>
<td>-.01299</td>
<td>.29870</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.05747</td>
<td>.74454</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td>.12051</td>
<td>.08485</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>-.25300</td>
<td>.12971</td>
</tr>
<tr>
<td></td>
<td>.22703</td>
<td>.46769</td>
</tr>
<tr>
<td>$t$</td>
<td>-.108</td>
<td>3.520</td>
</tr>
<tr>
<td>df</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.914</td>
<td>.001</td>
</tr>
</tbody>
</table>

Third grade students. The resulting analysis determined whether the mean value of the test scores for third grade students who took the CAB differed significantly from the mean value of the test scores of third grade students who took the CST. The
number of third grade students for this portion of the study totaled 81. Table 20 demonstrates the results of this analysis.

*Table 20: Mean Value of Third Grade Students*

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>3.05</td>
<td>2.84</td>
</tr>
<tr>
<td>CST</td>
<td>2.90</td>
<td>3.60</td>
</tr>
<tr>
<td>Difference</td>
<td>+.15</td>
<td>-.76</td>
</tr>
</tbody>
</table>

Table 20 shows third grade students who took the CAB and CST had minimal decrease in their CST scores in ELA (from 3.05 to 2.90, resulting in a difference of +.15) but a significant difference in the math portion of the test (from 2.84 to 3.60, resulting in a difference of -.76). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for the ELA portion of the test, but a significant difference for the math portion of the test. The results indicated that the mean CAB_{ela} and CST_{ela} (\( M = -1.14634, SD = .65020 \)) was significantly greater than the mean for CAB_{math} and CST_{math} (\( M = -.75610, SD = .80983 \)), \( t(81) = 8.455, p < .001 \). The standardized effect size index, \( d \), was .23 for CAB_{ela} and CST_{ela} and .94 for CAB_{math} and CST_{math} with considerable overlap in the distributions for the 5-point Likert ratings of CAB and CST. The 95% confidence interval for the mean difference between the CAB_{ela} and CST_{ela} was .0 and +.29 and CAB_{math} and CST_{math} was -.93 and -.58.
When focusing exclusively on the CAB score and CST score for third grade students, there are some notable results in the math test. To see if there were significant differences in third grade student scores on the CAB and the CST, the researcher used a paired samples $t$-test to evaluate within group comparisons. Table 21 shows the results of this analysis.

*Table 21: Paired Samples Test for Third Grade Students*

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAB ELA &amp; CST ELA</td>
<td>CAB Math &amp; CST Math</td>
</tr>
<tr>
<td>Mean</td>
<td>.14634</td>
<td>-.75610</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.65020</td>
<td>.80983</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td>.07180</td>
<td>.08943</td>
</tr>
<tr>
<td>Lower 95% Confidence Interval</td>
<td>.00348</td>
<td>-.93404</td>
</tr>
<tr>
<td>Upper 95% Confidence Interval</td>
<td>.28921</td>
<td>-.57816</td>
</tr>
<tr>
<td>$t$</td>
<td>2.038</td>
<td>-8.455</td>
</tr>
<tr>
<td>df</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.045</td>
<td>.000</td>
</tr>
</tbody>
</table>

**Fourth grade students.** The succeeding analysis determined whether the mean value of the test scores for fourth grade students who took the CAB differed significantly from the mean value of the test scores of fourth grade students who took the CST. The
number of fourth grade students for this portion of the study totaled 79. Table 22 shows the results of this analysis.

Table 22 shows fourth grade students who took the CAB and CST had minimal decrease in their CST scores in ELA (from 2.90 to 2.83, resulting in a difference of +.07) but a significant difference in the math portion of the test (from 3.08 to 3.26, resulting in a difference of -.18). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for both ELA and math.

Table 22: Mean Value of Fourth Grade Students

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>2.90</td>
<td>3.08</td>
</tr>
<tr>
<td>CST</td>
<td>2.83</td>
<td>3.26</td>
</tr>
<tr>
<td>Difference</td>
<td>+.07</td>
<td>-.18</td>
</tr>
</tbody>
</table>

When focusing exclusively on the CAB score and CST score for fourth grade students, there are some notable results in the math test. To see if there were significant differences in fourth student scores on the CAB and the CST, the researcher used a paired samples t-test to evaluate within group comparisons. Table 23 indicates the results of this analysis.
Table 23: Paired Samples Test for Fourth Grade Students

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAB ELA &amp;</td>
<td>CAB Math &amp;</td>
</tr>
<tr>
<td></td>
<td>CST ELA</td>
<td>CST Math</td>
</tr>
<tr>
<td>Mean</td>
<td>.07500</td>
<td>-.18750</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.75933</td>
<td>1.17024</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td>.08490</td>
<td>.13084</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>-.09398</td>
<td>-.44792</td>
</tr>
<tr>
<td>Upper</td>
<td>.24398</td>
<td>.07292</td>
</tr>
<tr>
<td>t</td>
<td>.883</td>
<td>-1.433</td>
</tr>
<tr>
<td>df</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.380</td>
<td>.156</td>
</tr>
</tbody>
</table>

Fifth grade students. The last analysis determined whether the mean value of the test scores for fifth grade students who took the CAB differed significantly from the mean value of the test scores of all fifth grade students who took the CST. The number of fifth grade students for this portion of the study totaled 76. Table 24 demonstrates the results of this analysis.

Table 24 shows fifth grade students who took the CAB and CST had minimal increase in their CST scores in ELA (from 2.51 to 2.82, resulting in a difference of -.31) but a significant increase in the math portion of the test (from 2.51 to 2.82, resulting in a
difference of -.31). Comparing the results between the CAB and CST for both subjects revealed there is no significant difference for both ELA and math.

When focusing exclusively on the CAB score and CST score for fifth students, there are some notable results in both the ELA and math tests. To see if there were significant differences in fifth grade student scores on the CAB and the CST, the researcher used a paired samples t-test to evaluate within group comparisons. Table 25 shows the results of this analysis.

*Table 24: Mean Value of Fifth Grade Students*

<table>
<thead>
<tr>
<th></th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB</td>
<td>2.51</td>
<td>2.51</td>
</tr>
<tr>
<td>CST</td>
<td>2.82</td>
<td>2.82</td>
</tr>
<tr>
<td>Difference</td>
<td>-.31</td>
<td>-.31</td>
</tr>
</tbody>
</table>

In summary, no significant difference in student scores on the CAB and CST in ELA was found for any groups of students. However, there was a significant difference in student scores on the CAB and CST in math for the following groups of students: All Students, African American, Asian, Non-EL, and Second and Third Grade Students. This portion of the study answers research question one (1a-1b).
Table 25: Paired Samples Test for Fifth Grade Students

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Pair 1</th>
<th>Pair 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAB ELA &amp;</td>
<td>CAB Math &amp;</td>
</tr>
<tr>
<td></td>
<td>CST ELA</td>
<td>CST Math</td>
</tr>
<tr>
<td>Mean</td>
<td>-.31169</td>
<td>-.31169</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.96327</td>
<td>.89236</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td>.10977</td>
<td>.10169</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>-.53032</td>
<td>-.51423</td>
</tr>
<tr>
<td>Upper</td>
<td>-.09305</td>
<td>-.10915</td>
</tr>
<tr>
<td>t</td>
<td>-2.839</td>
<td>-3.065</td>
</tr>
<tr>
<td>df</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Significance (2-tailed)</td>
<td>.006</td>
<td>.003</td>
</tr>
</tbody>
</table>

**Quantitative Study: Survey**

Two surveys were used in this study: teacher and principal. On the teacher survey, participants were asked to provide information on their experience with benchmark assessments. As Table 26 illustrates, many teachers tended to agree with each statement about benchmark assessments whereas very few teachers strongly disagreed with any of the statements about benchmark assessments. An extremely high percentage of teachers (94%) believed their students who did well on the CAB also did well on the CST in math. Furthermore, 100% of the teachers specified they had made changes to their instructional approach as a result of benchmark assessments; 83%
admitted they used benchmark assessments to plan and organize their lessons; 83% acknowledged they have changed the ways they work with specific groups of students as a result of benchmark assessments; as well as 94% believed benchmark assessments provided important diagnostic information on their students in math. Overall, benchmark assessments are viewed by teachers as an important tool for teaching and learning.

*Table 26: Teacher Survey Responses (All)*

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>My students who do well on the benchmark assessment in ELA generally do well on the California Standards Test in ELA.</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>My students who do well on the benchmark assessment in math generally do well on the CST in math.</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>I have made changes to my instructional approach as a result of benchmark assessments.</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>I use benchmark assessments to create activities for my students to do as an extension of my instruction.</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>My instructional approach is more focused on standards covered by benchmark assessments.</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>I now use benchmark assessments to plan and organize my lessons.</td>
<td>0</td>
<td>3</td>
<td>11</td>
<td>3</td>
</tr>
</tbody>
</table>

0% 0% 0% 0%
Table 26: Teacher Survey Responses
(Continued)

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Agree (3)</th>
<th>Strongly Agree (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark assessments play an important role in how I plan and organize small group instruction.</td>
<td>0%</td>
<td>24%</td>
<td>59%</td>
<td>18%</td>
</tr>
<tr>
<td>I use benchmark results to plan and organize differentiated lessons to meet my students’ needs.</td>
<td>0%</td>
<td>18%</td>
<td>59%</td>
<td>24%</td>
</tr>
<tr>
<td>Benchmark assessments have changed the ways I work with specific groups of students.</td>
<td>0%</td>
<td>18%</td>
<td>65%</td>
<td>18%</td>
</tr>
<tr>
<td>I spend more time working with specific groups of students than I have before because of benchmark assessments.</td>
<td>6%</td>
<td>35%</td>
<td>53%</td>
<td>6%</td>
</tr>
<tr>
<td>I now use a different approach to teaching specific groups of students based on benchmark results.</td>
<td>6%</td>
<td>29%</td>
<td>47%</td>
<td>18%</td>
</tr>
<tr>
<td>I use benchmark results in progress notices and report cards for my students in ELA and math.</td>
<td>0%</td>
<td>6%</td>
<td>61%</td>
<td>28%</td>
</tr>
<tr>
<td>The benchmark assessments provide important diagnostic information on my students in ELA.</td>
<td>0%</td>
<td>12%</td>
<td>59%</td>
<td>29%</td>
</tr>
<tr>
<td>The benchmark assessments provide important diagnostic information on my students in math.</td>
<td>0%</td>
<td>16%</td>
<td>65%</td>
<td>29%</td>
</tr>
</tbody>
</table>
The first two questions on the teacher survey responded to research question one with approximately 73% of the teachers either agreed or strongly agreed with the statement that their students who did well on the CAB also did well on the CST in ELA and 94% of the teachers for the math. The next three questions on the teacher survey responded to research question 2a with 100% of the teachers either agreed or strongly agreed they made changes to their instructional approach because of benchmark assessments, 94% of the teachers suggested they used benchmark assessments to create activities for their students to do as an extension of their instruction, and 80% of the teachers focused their instruction based on standards covered on benchmark assessments. The three questions following answered research question 2b with approximately 83% of the teachers either agreed or strongly agreed with using benchmark assessments to plan and organize their lessons, 77% of the teachers acknowledging benchmark assessments played an important role in how they plan or organize small group instruction, and 83% of the teachers using benchmark results to plan and organize differentiated lessons to meet their students’ needs. The three proceeding questions answered research question 2c which revealed approximately 83% of the teachers either agreed or strongly agreed that benchmark assessments had changed the ways they worked with specific groups of students, 59% of the teachers admitted spending more time working with specific groups of students because of benchmark assessments, and 65% of the teachers confirmed they used different approach to teaching specific groups of students based on benchmark results. Finally, the last three questions answered research question 2d, with
approximately 89% of the teachers admitting to using benchmark results in progress notices and report cards for their students, 88% of the teachers acknowledging benchmark assessments provided important diagnostic information on their students in ELA, while 94% of the teachers believe benchmark assessments provided important diagnostic information on their students in math.

As for the principal survey, only one of two principals completed the survey. Please refer to Appendix B for the questions on the survey. According to the principal’s response to the survey questions, she noted that she agreed with all statements on the survey with the exception that she did not use benchmark assessments and results in her Academic Conferences with her teachers. The following are a few of the statements she agreed with regarding benchmark assessments: benchmark assessments have influenced her leadership practices; it is an important part of meetings, teacher trainings, and Student Study Team (SST). Moreover, it has shaped her classroom observations in which she is more focused on teachers integrating benchmark assessments into the curriculum; using benchmark results to meet students’ specific needs; and focusing on the instructional rigor covered by benchmark assessments.

In summary, particular conclusions can be drawn from the quantitative portion of this study. Generally, both teachers and principals agree more often than disagree or have other opinions with each statement on the survey regarding benchmark assessments. First, they believe students who do well on the benchmark assessment in math also do well on the CST in math. Next, benchmark assessments have greatly influenced
instructional practices of teachers or leadership practices of principals. Many teachers acknowledge they have made changes to their instructional approach as a result of benchmark assessments along with using benchmark assessments to create activities for their students to do to extend on their instruction. Furthermore, their instructional approach is more focused on standards covered by benchmark assessments as benchmark assessments play an important role in how they plan and organize small group instruction. Finally, benchmark assessments play an important role in how teachers communicate student growth to parents via progress reports or report cards.

**Qualitative Study: Open-Ended Survey Questions & Interview**

**Open-ended survey questions.** Two open-ended survey questions were located at the end of both teacher and principal surveys. The first question asked, “*What thoughts do you have regarding the positive features of the benchmark assessments?*” Based on their responses, teachers believe benchmark assessments are very useful to teachers, students, and parents along with having noticeable impact on their weekly schedule due to the amount of time it takes to plan, teach, assess, and use the results to meet students’ academic needs via whole class or intervention as benchmark data are specific to each standard. Additionally, benchmark data provides essential information to share with parents per individual student reports along with teachers via class overview of each standard.

The second question asked, “*Are there any shortcomings as to the use of benchmark assessments?*” The following are responses to this question:
Some children do not try on standardized tests. I did not care enough about standardized tests as a K-12 student, so purposely did not try. The result being inaccurate information.

Some students seem to just mark bubbles on benchmarks rather than using their true knowledge; other assessment tools are vital to get the full picture.

While focusing on the benchmark, teachers need to place extra energy on the essential standards, while not excluding the other standards.

We are teaching how to bubble and test taking strategies rather than skills and strategies needed to pass the test on ability. If our benchmark was focused on the essential standards we should be asked to focus on these standards in our activities and instructions. Finally, instruction is not dependent on the state benchmark. It is a balance of multiple assessments; informal, formal, norm-, criterion referenced, and reading records. Our DIBELS and fluency assessments really drive our group structure in reading.

The benchmark should be closer aligned to the CST released questions.

Goes too wide and not deep for critical thinking skills; prefer Common Core.

If students have challenges outside of school, it could impact assessments and give a false reading of student capability.

ELLs don’t understand and have to be taught the skills.

Test taking skills need to be direct and throughout the year.
The writing task is included in the general standardized portion, and I don’t think subjective should be combined. I still think the writing task should be included, but not included in the overall score.

When asked, “What thoughts do you have regarding the positive features of the benchmark assessments,” the principal in this study responded that benchmark assessments and the CST are aligned. And when asked, “Are there any shortcomings as to the use of benchmark assessments,” she responded that multiple assessments are needed.

**Interview.** Twelve interview questions were asked of nine teachers. The first three questions answered research question 2a. The first question asked, “How have you changed your instructional approach as a result of benchmark assessments?” The responses provided by teachers indicated that benchmark assessments have forced them to “think outside the box” and make them aware of their own instructional weaknesses so they are required to reteach a standard differently from how they previously taught it. Also, they are more focused on essential standards that students performed poorly on. Moreover, they now examine and use data to determine what’s important to target instructionally and identify types of errors students make so they may teach students “tricks of testing.”

The second question asked, “What kinds of activities have you created as a result of benchmark assessments for your students to do as an extension of your instruction?” Respondents listed games, hands-on activities, matching games, group work, pocket chart
activities, whole unit of study, relevant materials, benchmark extension frames, and use of academic language as activities driven by benchmark assessment results.

The last question asked, “How is your instructional approach more focused on standards covered by benchmark assessments?” Teachers, in general, replied they were more instructionally focused on standards that most of their students were weak in that were covered on the benchmark assessments.

The next three questions answered research question 2b. The first of the three questions asked, “How do you use benchmark assessments to plan and organize your lessons?” Responses provided by teachers indicated benchmark assessments make them more focused on instructional pacing, teach standards the majority of the students struggle to learn, use benchmark questions in instruction and focus on instructional rigor, plan according to what standards are tested, and examine data to determine why students score low on particular standards.

The second question asked, “How do you use benchmark assessments to plan and organize small group instruction?” Respondents claimed they touched bases or worked with low achieving students; paired advanced learners with struggling learners, used supplemental materials to support student learning of particular standards, or go over benchmark questions to identify areas of student confusion and reasons.

The last question asked, “How do you use benchmark results to plan and organize differentiated lessons to meet your students’ needs?” Teachers indicated they used benchmark results to touch bases with low scoring students to identify questions students
did well on and why; which groups of students need what kind of support; plan accordingly for small groups or whole class; and cover questions that appear on the benchmarks using Board Math, Board ELA, or during workshop.

The subsequent three questions answer research questions 2c. The first question asked, “How have benchmark assessments changed the ways you work with specific groups of students?” Teachers interviewed admitted they were more focused on monitoring particular students who were struggling and inclined to work with these students in small groups and provide differentiated instruction.

The second question asked, “What part of your day do you spend working with specific groups of students as a result of benchmark assessments?” Teachers responded they did not spend a great amount of time working with specific groups of students based on benchmark assessment results for they only used a small portion of the day to work with students during workshop and tutoring, if they provided tutoring after school.

The last question asked, “How have benchmark results impacted your approach to teaching specific groups of students?” Responses indicated that benchmark results have helped teachers to see certain groups of students more clearly than they have in the past, that using other assessments is important for comparison and validity purposes, and identify holes that need to be filled in terms of what standards students struggle to achieve based on re-teaching.

The last three questions answered research question 2d. The first question asked, “How have you used benchmark results in progress notices and report cards for your
Responses provided by teachers varied. Some used only benchmark results for progress notices and report cards while others used benchmark results along with other sources of evaluation such as student work, project, or other assessments. Additionally, some teachers specified they are required by their principal to use benchmark results for progress notices and report cards.

The final two questions asked, “What diagnostic information do benchmark assessments provide on your students in ELA and math?” Teachers believed benchmark assessments provided them with a better focus on which standards are strengths and which are weaknesses/challenges, data to drive their instruction, allow them to see how students do on each standard, and information on which students need to be provided additional, differentiated instruction.

**Conclusion**

Many conclusions can be made about benchmark assessments based on both the quantitative and qualitative research methods used in this study. First, according to the comparison of CAB and CST scores, evidence shows on the ELA portion that although there were no significant differences between students’ CAB scores and CST scores, the results indicate that students tended to get the same result; therefore, conclusions can be made that the CAB has predictive validity with respect to the CST. Next, unlike the ELA portion, student scores on the math CAB and CST showed significant difference in six of ten subgroups (All Students, African American, Asian, Non-EL, Second Grade, and
Third Grade Students) (Refer to Table 27). In other words, students were more likely to score higher on the CST than the CAB; possibly indicating questions on the math CAB might be more difficult than math questions on the CST.

*Table 27: Significant Difference by Subgroups*

<table>
<thead>
<tr>
<th></th>
<th>CABela &amp; CSTela</th>
<th>CABmath &amp; CSTmath</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Students</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>African American</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Asian</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Latino</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Non-EL</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>EL</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Second Grade</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Third Grade</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Fourth Grade</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Fifth Grade</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

The significant difference in math scores between the CAB and the CST might be explained by other factors. For example, teachers could have been more comfortable, experienced, and trained in teaching mathematics in comparison to teaching English-language arts. A higher level of familiarity could have contributed directly to how well teachers taught mathematics, how much fidelity they put into teaching the subject, and
how they used the curriculum and other supplemental materials to support their instruction. Furthermore, teachers may have asserted more effort into the time they put into teaching math, interpreting the data, and using the data to provide follow-up instruction. Moreover, there are always questions about the accuracy of testing: environment, condition, and appropriateness. With respect to the significant differences identified, it is important to examine the structural frame of the organization (Bolman & Deal, 2008) to determine particular contributing factors to high student achievement in mathematics versus English-language arts on the CST.

In addition to comparing student scores on the CAB and CST, other conclusions can be drawn about benchmark assessments in general using teacher and principal surveys, which is reflective of Bolman and Deal’s (2008) human resource frame. Teachers and principals, as two stakeholders, play critical roles in student achievement. According to the surveys, benchmark assessments clearly impact teaching practices of teachers and instructional leadership practices of principals. Both groups surveyed agreed with the many benefits of benchmark assessments. The following are just a few of the benefits the participants agreed upon: Benchmark assessments play an important role in how I plan and organize small group instruction; Benchmark assessments have changed the ways I work with specific groups of students; I spend more time working with specific groups of students than I have before because of benchmark assessments; I now use a different approach to teaching specific groups of students based on benchmark
results; and I use benchmark results to plan and organize differentiated lessons to meet my students’ needs.

Along with comparing CAB and CST scores and teacher and principal surveys, open-ended survey questions and teacher interview also provided definite conclusions about benchmark assessments. Benchmark assessments provide essential data that transforms instruction. When looking ahead, teachers are able to strategically plan on how to use benchmark assessments to guide instruction. Additionally, teachers may use benchmark results to identify students’ strengths and weaknesses along with their own challenges with teaching particular standards. They are then able to group students to target particular standards students perform poorly on to ensure these students have a better opportunity for success on future benchmark assessments. All in all, although benchmark assessments is perceived by both teachers and administrators to be an important tool, few teachers know the full extent of the benefits of benchmark assessments and much professional development and teacher training is needed.
Chapter 5
SUMMARY AND CONCLUSIONS

Overview

Accountability in education puts incredible pressure on schools and districts nationwide to increase student achievement. This accountability pressure has resulted in an intense focus on using benchmark assessments to prepare students for the annual high stakes standardized tests. This study focused on whether there is a significant difference between student scores on benchmark assessments (CAB) and state standardized tests (CST) along with the role of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals. For the purpose of the study, student ELA and math scores on the CAB and CST were used as a measure to determine if there was significant difference between the two test scores. This chapter provides a summary of the study and its findings, a discussion of the research findings as they relate to information presented in the review of the literature, transformational leadership, data-driven decision making, policy implications, areas for future research, and conclusion.

Summary of Research

This study examined student scores in ELA and math on the CAB and CST to determine whether there is a significant difference between the two tests. Additionally, it explored the role of benchmark assessments on teaching and instructional leadership
practices in two Title I Program Improvement elementary schools in a Program Improvement California school district. The district has a total of 34 schools in Program Improvement for 2010-2011. The two schools in this study have been in Title I Program Improvement status for six years, the most in the district. They are two of six elementary schools in the district with this common designation. These two schools, amongst many in the district, with the exception of a few specialized schools, utilize benchmark assessments to improve student achievement on the annual state tests.

The researcher compared ELA and math student scores on the benchmark assessments and state tests from both schools in the study. After eliminating students who did not have both CAB and CST scores and special education students who took modified versions of the CST, the number of students used in this study totaled 317. The researcher used data from 2010-2011. The researcher also focused specifically on particular subgroups of students: African American, Asian, Latino, EL, non-EL; and second through fifth Grade. The findings of the study revealed noticeable patterns and statistically significant differences between student scores on the CAB and student outcomes on the CST for all 317 students along with particular groups of students: African American, Asian, non-EL, and second and third Grade. A summary of the findings is presented in the next section of this chapter.
Summary of Findings

While there were no significant differences between student scores for ELA and math on the benchmark assessments and state tests for particular subgroups (i.e. Latino, EL, and fourth and fifth grade students), statistically significant differences were found for all students, along with African American, Asian, non-EL, and second and third grade students for the math portion of the tests and interesting patterns emerged. In general, no significant difference was found between ELA student scores on the CAB and CST whereas significant differences were found between math student scores for both tests with six of the ten subgroups.

As suggested by the significant difference found between student scores on math benchmark assessments and state tests for six of the ten student subgroups, benchmark assessments have the potential to predict outcomes as found by other studies (Good, Simmons, & Kame’enui, 2001; Jacobsen, 2001; VanDerHeyden, Witt, Naquin; & Noell, 2001; Schilling et al., 2007; and McGlinchey & Hixson, 2004). Students in these groups were more inclined to score better on the CST than on the CAB. With that in mind, assumptions can be made that benchmark assessments are written at a more challenging level than state tests. On the other hand, although there was no significant difference found in ELA student scores, students tended to do fairly the same on both tests. Assumptions can be made that questions on the CAB and CST have a high positive correlation as they are closely aligned and it can be predicted that students will do about the same on both tests.
For example, as acknowledged by various teacher interview responses, many teachers recognized the importance of benchmark assessments in terms of providing immediate, ongoing targeted information pertaining to students’ academic strengths and weaknesses that teachers may employ to increase student learning (Chappuis et al., 2009; Coffey, 2009; CTB/McGraw-Hill, 2008, Noche, 2007; Olson, 2007; Sawchuk, 2009).

Likewise, the participants in this study also reported modifying their instruction as a result of benchmark assessments (Clune & White, 2008). The teachers reviewed problems from the assessments, aligned their instruction with the benchmark assessments, paid more attention to weak skills, focused on content that would be covered during each benchmark assessment period (Clune & White, 2008), or retaught content and skills assessed on benchmark assessments (Christman et al., 2009). The participants also used benchmark data to identify weaknesses in student performance in particular skills to reflect on their teaching practices (Olah et al., 2010). Furthermore, they used benchmark results to plan their instruction to focus specifically on areas students performed poorly on.

As previously annotated in chapter 2, research on how benchmark assessments are used by teachers in the classroom and by principals, as well as the impact of benchmark assessments on student performance is limited. Teachers and principals that were surveyed and interviewed in this study about their use of benchmark test results in instruction reported test results helped them monitor student progress and identify students’ skill gaps that led them to modify the curriculum and their instruction.
(Christman et al., 2009; Clune & White, 2008; Stecher et al., 2008). Furthermore, this study found that many of the participants confirmed that the benchmark test results helped them identify and modify gaps in curriculum and instruction (Stecher et al., 2008).

Discussion

The influence of benchmark testing on teaching practices of teachers and instructional leadership practices of principals will be discussed with reference to Bolman and Deal’s (2008) four organizational frames (structural, human resource, political, and symbolic). The structural frame focuses on the architecture of the organization, the design of its units and subunits, its rules and roles, its goals and policies; this frame emphasizes the importance in putting people in the right roles and relationships (Bolman & Deal, 2008). NCSD requires the use of CAB to prepare students in grades 2-8 in most schools in the district (with the exception of a few specialized schools) to take the high stakes CST near the end of the school year and achieve academic proficiency as a response to accountability required by the state. The goal of the school district could further be supported by principals who understand the importance of benchmark assessments and its relationship to the CST. To achieve greater investment in testing (CAB and CST) by principals, the district may want to consider making its vision and expectations crystal clear to its site administrators. In turn, leadership practices of principals will positively influence teaching practices of teachers, causing teachers to utilize benchmark assessments and results to plan instructional rigor that challenges all
students to make academic gains. The more informed the principals are, the better they are able to communicate expectations to teachers; thereby, allowing students to academically benefit.

Along with the structural frame, the human resource frame also plays an important role in benchmark testing and its influence on teaching practices of teachers and instructional leadership practices of principals. The human resource frame focuses on understanding people, their strengths and weaknesses, reason and emotion, desires and fears, and stresses dealing with issues by changing people (via training, rotation, promotion, or dismissal) (Bolman & Deal, 2008). Again, similar to the structural frame, all stakeholders must be considered. The district should make available and encourage principals to get trained on testing (CAB and CST), understand the importance of each test, and be cognizant on how to support teachers, who in turn will support student learning. The same can be said of teacher training on testing. The more teachers understand about the role of benchmark testing, how to integrate content standards into the curriculum, how to interpret and use benchmark results, the better they will be able to help students make the necessary academic gains needed to achieve proficiency throughout the grades. In summary, accountability comes with everyone understanding their role and responsibilities, utilizing their strengths to support their weaknesses, and appropriately supported in areas of stress and fear.

Equally important to the first two frames is the political frame, which views organizations as competitive and resource deficient, struggling for power and advantages
Like many schools and districts across the nation, NCSD’s goal is to increase student achievement on standards tests in order to avoid the negative impacts of NCLB. The two schools in this study are Title I, PI schools. Both are in low socioeconomic communities, populated with underserved culturally and linguistically diverse student populations, and have limited resources. Both schools are resource deficient and struggling to achieve academic proficiency each year.

The last of the four frames is the symbolic frame, which addresses issues of meaning and faith, with rituals, ceremonies, stories, plays, and cultures as central organizational life (Bolman & Deal, 2008). Both schools in this study are heavily impacted by testing. Year after year, the pressure for students to perform and achieve academic proficiency is evident. The goal of both schools is to get out of Program Improvement. In order to get out, the schools must either hit all of its subgroup AYP targets for two years running, or make significant overall gains for two consecutive years. The pressure is high for all parties involved: principals, teachers, and students. Currently and in the foreseeable future, the testing culture is central to the organizational life at both schools in this study as well as in schools throughout the district and across the nation.

**Implications**

This study on the influence of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals have many implications,
particularly in the area of transformational leadership, data-driven decision making, and policy analysis and action.

**Transformational Leadership**

According to results from teacher surveys and interviews, principals at both PI schools in this study are focused on increasing student achievement. Both expect teachers to teach standards covered on benchmark assessments with the greatest emphasis placed on essential standards (highest level of frequency on the test). However, some teachers see “teaching to the test” as depriving students of greater opportunities that teaching for social justice provides and is necessary in order for students to function and contribute constructively to society. As transformational leaders, principals must make themselves and others more aware of the importance and value of the work of increasing student achievement should not be based on “teaching to the test,” rather, “…to transcend self-interest for the sake of the organization” (Yukl, 2006, p. 280).

In order to become a nation that is exceptional and lead in a global economy unmatched by any other in the world, the quality of the United States education system must go beyond preparing students to achieve proficiency on an exam. After all, education is a powerful tool for reducing poverty and inequality along with setting a foundation for creating and sustaining economic growth and opportunities (The World Bank Group, 2011). Education has the great potential to create productive critical thinkers who are efficient, pliable, and collaborative workers along with expanding their
aptitude to successfully complete tasks or use industrious technologies (Decker, 1997; Radcliffe, 2009).

Additionally, transformational leaders must understand that when trying to generate adaptive change, they must work with “…differences, passions, and conflicts in a way that diminishes their destructive potential and constructively harnesses their energy” (Yukl, 2006, p. 102). Leadership is about influencing and facilitating; it is “…influencing others to understand and agree about what needs to be done and how to do it, and the process of facilitating individual and collective efforts to accomplish shared objectives” (Yukl, 2006, p. 10). Many factors determine a leader’s level of influence. First and foremost, the leader must be knowledgeable. With knowledge comes credibility. Next, the leader must be able to effectively communicate her vision along with respecting and valuing the vision of her staff. On top of that, the leader must be collaborative with all stakeholders in her efforts to achieve a particular outcome. After that, the leader must demonstrate reliable follow-through and provide appropriate support as necessary.

Transformational leaders must also know how to “orchestrate the conflict” to address the issue of what needs to get done to get students to achieve the standards (Heifetz & Linsky, 2002). To “orchestrate the conflict,” transformational leaders should “create a holding environment” where all stakeholders can tackle tough issue of addressing student achievement on standards tests (Heifetz & Linsky, 2002). Just like any other tough issue, transformational leaders should also know how to “control the
temperature” so the discussion stays within a productive range. Heifetz and Linsky (2002) suggested various ways of controlling the heat during the course of discussion.

To raise the temperature, transformational leaders can do the following: 1) Draw attention to the tough questions, 2) Give people more responsibility than they are comfortable with, 3) Bring conflicts to the surface, and 4) Protect gadflies and oddballs (Heifetz & Linsky, 2002, p. 111). To lower the temperature, transformational leaders can do the following: 1) Address the technical aspects of the problem, 2) Establish a structure for the problem-solving process by breaking the problem into parts and creating time frames, decision rules, and clear role assignments, 3) Temporarily reclaim responsibility for the tough issues, 4) Employ work avoidance mechanisms, and 5) Slow down the process of challenging norms and expectations (Heifetz & Linsky, 2002, p. 111). Once the temperature has been controlled, transformational leaders need to set the “pace of the work” because change in anything (behavior, practices, habits, etc.) requires time. Rushing the work would be counterproductive. In short, transformational leaders must be skilled communicators and emotion regulators.

Finally, transformational leaders should show the stakeholders how the change will impact the future of the organization to sustain their motivation and drive to strive. As Yukl commented, “Before people will support radical change, they need to have a vision of a better future that is attractive enough to justify the sacrifices and hardships the change will require” (2006, p. 316). It is critical for transformational leaders to see their vision from beginning to end, have an action plan, and effectively articulate their action
plan to all stakeholders. It is also important that all parties involved understand their role and responsibilities along with accountability measures the leader will take to ensure the change takes place. Moreover, the leader must be prepared to provide appropriate support and resources.

**Data-Driven Decision Making**

Both schools in this study are data-driven schools as both are in Program Improvement Year 5. They use ELA and math benchmark assessments and data to assist them in preparing students to score proficiency on the end-of-year annual state tests in addition to making the necessary strides toward improving learning for all students. NCSD has acknowledged that in order to ensure that every student by name is successful the district must continue to focus on the achievement rates of students not meeting the state standards. There are numerous quantitative and qualitative assessments the district uses to measure student achievement. The following is a list of all assessments used by the district: Standardized Testing and Reporting system (STAR), NCSD benchmark assessments, California English Language Development Tests (CELDT), English Language Development Quarterly Assessments, suspension and expulsion data, student attendance data, staff attendance data, Healthy Kids Survey, school climate surveys, and California High School Exit Exam (CHSEE).

**Data-driven practices.** Although the district provides a spectrum of assessments to measure student achievement, both schools in this study have placed major emphasis specifically on two assessments previously mentioned, the STAR and SCUSD
benchmark assessments and two not mentioned, Reading Lion Assessments and Saxon Math Assessments (supported by the district’s Elementary Curriculum department through professional development and training specialists). Schools throughout the district support the use of data to drive instruction via numerous opportunities for teachers to examine data from all four assessments during common planning time (CPT), curriculum meeting, staff meeting, and professional development. During these times of opportunity, the administrator, training specialist, or district employee(s) lead teachers in examining data to determine areas of focus, instructional and engagement strategies, and targeted students. Teachers work collaboratively in grade level groups to determine the most appropriate approach and share ideas on how to achieve greater success by the next determined assessment date.

**Suggestions for building a data-drive culture.** There are some suggestions that schools in this study should take into consideration if it is to build a data-driven culture that will lead to improved learning for all students. First, teachers can identify better with other teachers with experiences relevant to their own. Having a classroom teacher who’s experienced with using data to drive his/her instruction facilitate a training is by far more beneficial to other teachers than to have someone who’s not as connected with the classroom such as an administrator or training specialist facilitate a training, especially if he/she has been out of the classroom for a while. Next, because examining data can be very complex, it would be helpful to create a concrete step-by-step visual in which the participants and facilitator examine the same data to determine target group, areas of
focus, and strategies to use to engage students in the learning process. Then, once facilitator and participants have examined the same data and identified areas of focus, the participants should work in grade level teams to examine their own data and determine what outcome they would like to see. Finally, each grade level group should create a visual presentation to share with other grade levels so they can benefit from one another’s perspectives and get feedback from their peers. All in all, having a plan that is simple to follow and well-thought out can result in building a data-drive culture that will lead to improved learning for all students.

Policy Analysis & Action

This study, like other research on this topic, has many policy implications at the local, state, and national levels. As previously stated, the United States is a nation that strives to be exceptional and lead in a global economy unmatched by any other. This demand to be the best has resulted in the nation’s intense focus on quality education for all children in this country. Education plays an important role in creating productive, efficient, pliable, and collaborative workers and critical thinkers as well as expanding a worker’s aptitude to successfully complete tasks or use industrious technologies (Decker, 1997; Radcliffe, 2009). According to the U.S. Department of Labor (1990), increases in educational opportunities were responsible for 20% growth in worker productivity in the United States in recent decades. In short, educational leaders must make informed decisions based on research and data. Moreover, policies must reflect the local, state, and national’s efforts to prepare career or college ready students.
Considering existing research supports the importance of benchmark assessments in classrooms due to its positive possibilities with respect to both policy and pedagogy (Gorlewski, 2008), an important policy implication is teacher training on the use of benchmark data. According to surveys and interviews used in this study, benchmark data are not being used by teachers to understand the extent to which students really understand content (Nabors Olah et al., 2010). This study revealed that both teachers and principal demonstrate limited knowledge in evaluating the quality of assessment data (Bursch et al, 2010) as well as lack the knowledge, resources, and support to connect assessment results to teaching and learning (Datnow et al., 2007; Kerr et al., 2006; & Young, 2006).

This policy implication is an important component of student achievement. The purpose of testing is data. What is data if its benefits are not fully understood? Likewise, teacher acceptance of data usage is low for those who have not been trained, prepared, or supported well in using assessment data to modify instruction as they often become frustrated and resistance with analyzing student data (Trimble et al., 2005). Furthermore, Blanc, Christman, Liu, Mitchell, Travers, and Bulkley (2010) suggested “…student learning will not improve unless attention is paid to developing strong school leaders who promote data-driven decision making within a school culture focused on strengthening instruction, professional learning, and collective responsibility for student success” and that “…a well-designed system of interim assessments is dependent on the knowledge
and skills of the school leaders and teachers who are responsible for bringing the system
to life in schools” (Blanc et al., 2010, p. 206).

**Areas for Future Research**

This study contributes to the body of research regarding the relationship between
benchmark assessments and state tests. This research is very current and relevant as
many schools and districts nationwide are administering periodic benchmark assessments
to complement the end-of-year state testing and increase student learning as a response to
accountability pressures to raise student achievement despite the lack of research base
concerning its effectiveness (Hender
son et al., 2007). There are several areas for future
study related to this topic.

First, this study, like many, included a small sample of teachers, principals, and
schools in one school district. Future studies on this topic should include larger numbers
of teachers, principals, and schools to control for possible school and student
characteristics. On the contrary, future studies could have a narrower focus by studying a
particular group of teachers and their classroom interventions used in conjunction with
benchmark assessments or the types of classroom activities the teachers develop as well
as particular principals and their knowledge of and role in benchmark assessments.

Next, aggregated results from benchmark assessments provide a wealth of
information for school administrators (Olson, 2007). Future studies should survey and
interview school board members, the superintendent, and district administrative staff due
to their role in setting policy, adopting programs, and organizing the delivery of instructional services to meet the needs of children (Olson, 2007).

Last, it is also important for future research on benchmark assessments to consider the impact of teacher training on using benchmark results as teacher knowledge is one of the most influence factors in students’ success in school (Darling-Hammond, 2000). Do a teacher survey on how they prepare students for benchmark assessments and the CST, the role of benchmark assessments in their daily schedule and yearlong plan, and their usage of benchmark results.

**Conclusion**

It is the responsibility of all stakeholders in education to prepare all students to be career or college ready. This hefty task should start as early as elementary school. Students are required to take the state tests as early as second grade and students in first grade are already getting their feet wet with taking benchmark assessments, which help prepares students in grades three to eleven, for the annual end-of-year state tests.

This research focused on the relationship between benchmark assessments and state tests. Specifically, it focused on whether student scores on the CAB are significantly different from student scores on the CST. Benchmark assessments are presumed to help increase students’ academic achievement on statewide tests despite the lack of research base concerning its effectiveness (Henderson et al., 2007). Up until this research, no study had been conducted to investigate the relationship between student
scores on the final benchmark assessments and student scores on the CSTs in Title I Program Improvement elementary schools. Furthermore, no study exists on the influence of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals in these low achieving elementary schools.

The results of this study showed the link between benchmark assessments and state tests along with the importance of benchmark assessments on teaching practices of teachers and instructional leadership practices of principals. Based on this study, conclusions can be made that with appropriate teacher and administration training on implementing benchmark assessments, interpreting benchmark results, providing appropriate intervention to students based on their results, resource allocation to support student achievement, and consistent administrative follow-up could result in improved student learning and outcomes, as well as students who are career or college ready.
APPENDICES
Appendix A
Teacher Survey

1. I teach at the following school:
   - □ Ethel Phillips
   - □ Freeport
   - □ Jedediah Smith
   - □ John Still K-6
   - □ Oakridge
   - □ Pacific

2. I currently teach the following grade level:
   - □ 2\textsuperscript{nd} Grade
   - □ 3\textsuperscript{rd} Grade
   - □ 4\textsuperscript{th} Grade
   - □ 5\textsuperscript{th} Grade
   - □ 6\textsuperscript{th} Grade
   - □ Combination Class ____________________________

3. I have taught my grade level for ___ years.
   - □ 0-1 year
   - □ 2-5 years
   - □ 6-10 years
   - □ 11+ years
4. I have taught a total of ___ years.

- [ ] 0-1 year
- [ ] 2-5 years
- [ ] 6-10 years
- [ ] 11+ years

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>My students who do well on the benchmark assessment in ELA generally do well on the California Standards Test in ELA.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>My students who do well on the benchmark assessment in math generally do well on the CST in math.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>I have made changes to my instructional approach as a result of benchmark assessments.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>I use benchmark assessments to create activities for my students to do as an extension of my instruction.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>My instructional approach is more focused on standards covered by benchmark assessments.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I now use benchmark assessments to plan and organize my lessons.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Benchmark assessments play an important role in how I plan and organize small group instruction.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I use benchmark results to plan and organize differentiated lessons to meet my students’ needs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Benchmark assessments have changed the ways I work with specific groups of students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I spend more time working with specific groups of students than I have before because of benchmark assessments.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I now use a different approach to teaching specific groups of students based on benchmark results.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I use benchmark results in progress notices and report cards for my students in ELA and math.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>The benchmark assessments provide important diagnostic information on my students in ELA.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
14. The benchmark assessments provide important diagnostic information on my students in math.

15. What thoughts do you have regarding the positive features of the benchmark assessments?

16. Are there any shortcomings as to the use of benchmark assessments?
Appendix B

Principal Survey

1. I have been principal at my current school for ___ years.
   - [ ] 0-1 year
   - [ ] 2-5 years
   - [ ] 6-10 years
   - [ ] 11+ years

2. I have been principal for a total of ___ years.
   - [ ] 0-1 year
   - [ ] 2-5 years
   - [ ] 6-10 years
   - [ ] 11+ years

3. I have had prior experience implementing benchmark assessments as a school administrator.
   - [ ] Yes
   - [ ] No

4. I have had prior experience implementing benchmark assessments as a teacher.
   - [ ] Yes
   - [ ] No
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Benchmark assessments have influenced my leadership practice.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Academic Conferences are more focused on benchmark assessments and results.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>SSTs and 504 plans are more focused on benchmark results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Curriculum Meetings are more focused on benchmark results.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>My classroom observations are focused on the integration of benchmark assessments into the curriculum.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>My classroom observations are focused on the use of benchmark results to meet students’ specific needs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>My classroom observations are focused on standards covered in benchmark assessments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>My classroom observations are focused on the level of instructional rigor covered by benchmark assessments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>I hold meetings and trainings on how teachers can use benchmark assessments in their lesson planning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>I hold meetings and trainings on how teachers can use benchmark results to differentiate their instruction.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>My approach to working with teachers has changed as a result of benchmark assessments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12. My instructional discussion with teachers is more focused on standards covered in benchmark assessments.

13. My assessment discussion with teachers is more focused on benchmark assessments.

12. What thoughts do you have regarding the positive features of the benchmark assessments?

13. Are there any shortcomings as to the use of benchmark assessments?
Appendix C

Teacher Interview

1. How have you changed your instructional approach as a result of benchmark assessments? RQ2a

2. What kinds of activities have you created as a result of benchmark assessments for your students to do as an extension of your instruction? RQ2a

3. How is your instructional approach more focused on standards covered by benchmark assessments? RQ2a

4. How do you use benchmark assessments to plan and organize your lessons? RQ2b

5. How do you use benchmark assessments to plan and organize small group instruction? RQ2b

6. How do you use benchmark results to plan and organize differentiated lessons to meet your students’ needs? RQ2b

7. How have benchmark assessments changed the ways you work with specific groups of students? RQ2c

8. What part of your day do you spend working with specific groups of students as a result of benchmark assessments? RQ2c

9. How have benchmark results impacted your approach to teaching specific groups of students? RQ2c

10. How have you used benchmark results in progress notices and report cards for your students in ELA and math? RQ2d
11. What diagnostic information do benchmark assessments provide on your students in ELA? RQ2d

12. What diagnostic information do benchmark assessments provide on your students in math? RQ2d
Appendix D

Principal Interview

1. To what degree has benchmark assessments influenced your leadership practice? RQ3a

2. To what extent do benchmark assessments and results influence how you conduct Academic Conferences? RQ3a

3. In what ways are benchmark results used during SST and 504 Plan meetings? RQ3a

4. How are benchmark results used during Curriculum Meetings? RQ3a

5. How have benchmark assessments influenced your classroom observations? RQ3b

6. How have benchmark results influenced your classroom observations with regards to meeting students’ specific needs? RQ3b

7. In what ways have standards covered in benchmark assessments been a part of your classroom observations? RQ3b

8. What role does the level of instructional rigor covered by benchmark assessments play in your classroom observations? RQ3b

9. To what degree do benchmark assessments in lesson planning influence how you hold meetings and trainings? RQ3c

10. In what ways have using benchmark results to differentiate instruction impacted how you hold meetings and trainings? RQ3c
11. How has your approach to working with teachers changed as a result of benchmark assessments? RQ3c

12. How much of your instructional discussion with teachers been focused on standards covered in benchmark assessments? RQ3c

13. How much of your assessment discussion with teachers been focused on benchmark assessments? RQ3c
REFERENCES

http://www.accellos.com/curriculumassociates

Amrein-Beardsley, A. (2009). The unintended, pernicious consequences of “Staying the
course” on the United States’ No Child Left Behind Policy. International Journal
of Education Policy and Leadership, 4(6). Retrieved on December 30, 2011 from

second, and third degree: Educators’ responses to high stakes testing. Education
Policy Analysis Archives: A Peer-Reviewed, Independent, Open Access,
Multilingual Journal, 18(14). Retrieved December 29, 2011 from


Education Assessment Evaluation Accountability, 22, 53-72.

VA: Center for Public Education. Retrieved August 15, 2011 from
http://www.centerforpubliceducation.org/Main-Menu/Instruction/High-stakes-
testing-and-effects-on-instruction-At-a-glance/High-stakes-testing-and-effects-on-
instruction-PDF-.PDF


http://www.acsi.org/LinkClick.aspx?fileticket=sEXRSzxGNLA%3D&tabid=934


from
http://www.leadered.com/06Symposium/pdf/USES%20OF%20BENCHMARK%20TESTS.pdf

Springboard Schools (2006). Bringing the state and locals together: developing effective
data systems in California School Districts. A Network Committed to
Achievement, Equity, & Inquiry.

Stecher, B., Epstein, S., Hamilton, L.S., Marsh, J.A., Robyn, A., McCombs, J.S., Russell,

close achievement gaps. Theory into Practice. 11-18.


imperative for educational reform. A Report to the Nation and the Secretary of


Tingey, R. A. (2009). High-stakes testing under the No Child Left Behind (NCLB) Act: How has it impacted school culture? ProQuest LLC, Brigham Young University.


Retrieved August 17, 2011 from

http://ed.gov/parents/academic/involve/nclbguide/parentsguide.pdf


http://dpi.state.wi.us/oea/pdf/bas.pdf