BEATING THE HIGH STAKES TESTING GAME: A THREE-YEAR STUDY OF IMPROVEMENT RATES ON THE TAKS SOCIAL STUDIES EXIT EXAM

Barbara Anne Evans, B.A., M.A.T.

Dissertation Prepared for the Degree of

DOCTOR OF EDUCATION

UNIVERSITY OF NORTH TEXAS

May 2006

APPROVED:

Alexandra Leavell, Major Professor
Tandra Tyler-Wood, Minor Professor
Ronald Wilhelm, Committee Member
James D. Laney, Program Coordinator
Mary Harris, Interim Chair of the Department of Teacher Education and Administration
M. Jean Keller, Dean of the College of Education
Sandra L. Terrell, Dean of the Robert B. Toulouse School of Graduate Studies
Evans, Barbara Anne, *Beating the high stakes testing game: A three-year study of improvement rates on the TAKS social studies exit exam*. Doctor of Education (Curriculum and Instruction), May 2006, 136 pp., 7 tables, 2 figures, references, 252 titles.

The Texas high school class of 2005 faced a defining test that had no precedent in Texas and little nationally. Social studies testing is a relatively new addition to the world of high stakes testing currently impacting United States high schools. Although other diploma dependent areas of mandated testing have some testing history and, therefore, related paradigms for curriculum and instructional assistance, the area of social studies largely lacks that perspective. Texas Education agency provided specific school grant monies and training for the purpose of preparation for the social studies exams. This quasi-experimental study examines the scores to learn whether or not any statistically significant differences in social studies scores would exist between the schools that participated in the TEKS/Tools Training Program and the schools that did not participate in the TEKS/Tools Training Program. The two primary at-risk groups in Texas, Hispanic and low SES, were analyzed for statistically significant differences in scores. Independent t tests and ANCOVA were used to analyze the score differences between program schools and non-program schools. Results relate to individual school staffing and implementation. The at-risk groups remained flat in score gains whether they were part of the program schools or not. Results relate to differences in learning and teaching for at risk groups. A separate trend analysis was used on the program target school which was the only school with three years of scores to determine improvement from grade 9 to 10 to 11 on the social studies TAKS test scores. Results from the repeated measures analysis indicated a statistically significant linear trend in the program target school’s TAKS social studies mean gain scores across the 9th, 10th, and 11th grade levels.
ACKNOWLEDGEMENTS

The author extends a grateful and enduring acknowledgement to my students. I have learned from and been energized by you. The privilege of being allowed to guide you has been a gift that has given life meaning.

I wish to thank my brilliant committee: Dr. Alexandra Leavell for wit, insight, understanding, patience, and so much more; Dr. Tandra Tyler-Wood for connecting research to practice and truly understanding students as individuals; and Dr. Ronald Wilhelm for steady guidance, insistence on quality, and the soul of a reformer. I could not have had a more interesting and invigorating academic experience; my committee was the core reason.

Acknowledgement and appreciation go to Harishkumar Kedarnath, Dr. Randall E. Schumacker, Dr. Mike Clark, and the late Dr. L. Fred Thomas for your collective contribution to my discernment of statistics. Without your focus, I could not have developed awareness and appreciation of data driven decision-making in education.

I wish to acknowledge the gracious help of Dr. J. D. Kennedy and Dr. Roxanne Wilson. You gave me access to the tools and school settings to make this study authentic and meaningful. In tandem with Texas Education Agency, you empowered and enabled the stages of this study. My appreciation of supportive leadership like yours is immeasurable.

Finally, I wish to thank my family. Richard Chowning and Ashlye Stewart, you are my inspiration and greatest support. Your encouragement was consistent and overwhelming; I could not have continued without you.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Background and Importance of Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Program Development</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Purpose of the Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Method</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Research Questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Problems as Null Hypotheses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limitations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Definitions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assumptions</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>REVIEW OF LITERATURE</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Case for Standards</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A Recent History of High Stakes Testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Focus on Social Studies Standards and Assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>History of Accountability in Texas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Texas Assessment of Knowledge and Skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Issues and Concerns Regarding High Stakes Testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lack of Research and Quality of Research</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>RESEARCH DESIGN</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Overview, Statement of the Problem, Design</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participants: The Decatur Sample School</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Description of the Program and Comparison Groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Procedures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Description of Data Sources</td>
<td></td>
</tr>
</tbody>
</table>
Testing Procedures
Collection of Data
Decatur Sample Group Analyses
Campus Comparison Group Analyses
Statement of the Research Questions & Hypothesis
Summary

4. RESULTS ............................................................................................................... 67
   Introduction
   TAKS Social Studies Gain Score Analysis
   Socioeconomic Status Analysis
   Hispanic Subpopulation Analysis
   Decatur Sample Trend Analysis

5. CONCLUSIONS AND RECOMMENDATIONS ............................................. 73
   Discussion and Conclusions
   Educational Importance
   Recommendations

Appendices

   A. SOCIAL STUDIES HIGH STAKES TESTING BY STATE ..................... 93
   B. TEKS/TOOLS GRANT INFORMATION .............................................. 95
   C. HOW CAMPUS COMPARISON GROUPS FORMED ............................. 100
   D. TEST DEVELOPMENT PROCESS ..................................................... 113
   E. GAIN SCORES – CAMPUS COMPARISON GROUP SCHOOLS .......... 117

REFERENCES ......................................................................................................... 119
## LIST OF TABLES AND FIGURES

### Tables

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>TAKS Social Studies Gain Score T Test Analysis</td>
<td>69</td>
</tr>
<tr>
<td>2.</td>
<td>Original and Adjusted SES TAKS Social Studies Means</td>
<td>71</td>
</tr>
<tr>
<td>3.</td>
<td>Analysis of Covariance on Low SES TAKS Social Studies Scores</td>
<td>71</td>
</tr>
<tr>
<td>4.</td>
<td>Original and Adjusted Hispanic TAKS Social Studies Means</td>
<td>73</td>
</tr>
<tr>
<td>5.</td>
<td>Analysis of Covariance on Hispanic TAKS Social Studies Scores</td>
<td>73</td>
</tr>
<tr>
<td>6.</td>
<td>Decatur Sample Mean Gain Scores for 9th, 10th, 11th Grades</td>
<td>74</td>
</tr>
<tr>
<td>7.</td>
<td>Repeated Measures for 9th, 10th, 11th Grades (Decatur Sample)</td>
<td>75</td>
</tr>
</tbody>
</table>

### Figures

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Research and Study Timeline</td>
<td>42</td>
</tr>
<tr>
<td>2.</td>
<td>Decatur Campus Comparison Group</td>
<td>49</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Background

The current emphasis on standardized tests to measure the quality of education can be traced back to September 1989, when President George H. W. Bush joined the nation's 50 governors to convene an Education Summit in Charlottesville, VA. One of the outcomes of that meeting was the establishment of the National Education Goals Panel. Charged with issuing annual reports on state and national progress toward education goals, governors held eight regional public hearings to get feedback from educators and the public to determine the indicators of such progress. Thus, the standards movement was born.

In March 1994, President Bill Clinton signed the Goals 2000 Educate America Act, codifying eight national education goals. Assistant Secretary of Education Diane Ravitch affirmed his initiative as a positive move toward improving the effectiveness of American education: "Standards can improve achievement by clearly defining what is to be taught and what kind of performance is expected" (Ravitch, 1995; p. 25). In 1996, the Goals Panel's annual report reified the standards movement for local and state educators who were charged with increasing student achievement and designing new forms of assessment to measure that achievement. Not surprisingly, American education has become increasingly focused on standards as markers for quality in education. (IASA, Goals 2000).

In January 2001, George W. Bush announced his plan for improving American education. Less than a year later, the No Child Left Behind (NCLB) Act of 2001 became federal law (US Department of Education, 2004). NCLB purported to support greater choice for parents and students, particularly those attending low-performing schools; more flexibility for states and
local educational agencies in the use of federal education dollars; a stronger emphasis on reading, especially for our youngest children; and increased accountability for states, school districts, and schools. Although a few argued with the intent of the law, the implementation of NCLB has proven to be controversial. Regardless of one’s opinion, the presence of this mandate has had an inarguable affect on accountability required by states.

The demand for accountability measures to report adequate yearly progress has increased the use of standardized tests exponentially as indicators of successful achievement. New York and Texas are at the apex of rigor for both standards and testing (Pahl, 2003). Both states use standardized tests as gatekeepers for student advancement. Even though the high stakes label has traditionally been limited to reading and mathematics testing, other academic areas are now receiving equal attention under the application of standards.

Social studies has not typically been an academic area subject to standardized testing. Indeed, in some states, social studies instruction has been marginalized because it is not part of the test battery used for accountability purposes (Manzo, 2005). Such shortsightedness is no longer the case in Texas. In 1999, with the passage of SB 103, social studies was added as an assessment area for the Texas Assessment of Knowledge and Skills, the state test used for accountability, making it one of nine states (see Appendix A) to include social studies standardized state exams (Center of Educational Policy, 2003). As noted by the late U.S. Senator Paul Wellstone (2000) “High stakes tests are a harsh agenda for America’s children.” Clearly, the Texas high school class of 2005 faced that harsh reality as their graduation became contingent upon passing all subtests of the TAKS.

Suddenly, Texas social studies teachers were under pressure to prepare children for a test, purported to be more academically rigorous than its predecessor, in a content area not previously
represented on the state test. In an effort to be proactive, as no preparation materials were immediately provided by the state, Decatur ISD social studies educators sought local, state and national models they could emulate in order to provide excellent instruction and assure students would be ready for the test. Although there existed some evidence of rigorous upgrades to curriculum by certain states (Gaylord, 2003), nothing was readily available that would transfer neatly into Decatur schools.

Program Development

Decatur Independent School District proceeded to design and implement an instructional program crafted to uphold the content standards set forth by the TEKS and assessed by the TAKS. Decatur applied for grant funding; addressed the alignment of the TEKS and the TAKS; assessed student needs, with particular concern for the subgroups considered at risk; redesigned the social studies curriculum; trained teachers in strategies and up to date delivery systems; and implemented a pilot study that became useful in continued formative program design and analysis.

To support the endeavor, Decatur Independent School District applied for and was awarded a TEKS/ Tools Grant from the Texas Education Agency (Social Studies Center, 2001). Decatur became the target school for the purpose of this study. As described above, the adaptation of the curriculum and the resultant changes in teacher practice might appear to be a natural evolution in response to the threat of a high stakes TAKS Social Studies test; however, without the state incentive of the TEKS/Tools Grant Program it is very possible such changes probably might not have occurred.
The grant money allowed for (among other things) hiring substitutes so teachers could not only meet and plan collaboratively to redesign curricula, but also participate in an extensive professional development program in the appropriate use of the TEKS curriculum and intervention strategies. The allocated funds provided time, training, and support, which allowed DISD to rebuild the social studies curriculum as well as to refine the delivery system with staff development and training tailored to the demands of conceptually more rigorous student testing.

Knowing the existing math and reading TAKS tests were tightly aligned with the corresponding TEKS; Decatur began with the TEKS as their foundation for the needs assessment. Participating teachers began a total deconstruction of the TEKS with attention to scope and sequence of knowledge and consideration of appropriate skills for each level. The strategies employed in the Decatur TEKS/Tools Grant Program realignment (see Appendix B) were grounded in the theories of Bloom, Vygotsky, Piaget, Bruner, and other constructivist and cognitive-learning theorists (see chapter 3 for a complete description).

A full assessment of students’ needs and lack of knowledge and/or skills in the various curricular areas ensued. There was particular concern about student subgroups traditionally considered at-risk for academic failure (McNeil & Valenzuela, 2001). Texas population in those groups is high (Standard & Poors, 2005). Decatur ISD, the target district for this study, reflects that emerging demographic trend toward higher numbers of at-risk children.

The Decatur program design proceeded as a formative process. With the intent to create a picture of the whole learner, Decatur teachers gathered student data on general prior knowledge, specific social studies content knowledge and skills, reading ability, learning styles and differences, and general academic history. The synthesis of that information helped to create a profile for each student, which served as a guideline for individual class placement and
instruction. Equally important, the data collected helped affect overall program design, curricular changes (vertical and horizontal alignment of the social studies curriculum), and thus, teacher professional development and lesson planning. The curricular alignment process and intervention strategies for social studies teaching that emerged constitute a progressive integrative approach, also referred to as a reconceptualized approach by researchers who specialize in the social studies learning environment (Segall, 2002).

Teacher training was a critical facet of the program design. The related teacher training aspects of the approach follow the National Staff Development Council (NSDC) guidelines for quality professional development to address the context, content, and processes needed to close the gap between national and state instructional standards and student achievement when the evaluation for mastery on those standards is a high stakes test (NSDC, 2002-2004). The professional development aspect of the TEKS/Tools Grant emphasized the close alignment of the Texas Essential Knowledge and Skills (the state social studies curriculum guidelines) to the TAKS test objectives (Social Studies Center, 2001). The program goal involved enhancement of teachers’ knowledge and implementation of the TEKS social studies curricula as a vehicle for Texas Assessment of Knowledge and Skills or TAKS preparation (Segall, 2001). The goal of teaching a rich, rigorous, aligned curriculum strongly informed the student preparation method for success on the TAKS (Social Studies Center, 2000-2001).

The reconceptualized view of social studies teaching is a field of emerging and innovative ideas that stresses the inclusion of inter- and cross disciplinary concepts, of culture and language as historical representation, and of approaching social studies in a manner that encourages critical mindedness (Berkhofer, 1995). One example of practical implication would be the active use of primary documents. Whenever instructors use discourse and debate, the
practice of using primary documents forms part of a rigorous and rich curriculum (Wineburg, 1991, and Jenkins, 1991). This strategy stimulates critical thinking with the possibility of the student transferring that skill to future test questions.

The program was implemented beginning with a pilot study using the Decatur students. Change is never easy and the success of any change or reform hinges upon how it is implemented and the effects of that implementation (Goodson & Foote, 2001). Plans necessarily change and thus the assumptions inherent in the Concerns-Based Adoption Model (C-BAM) (Hall and Hord, 2001), especially the levels of use or implementation, provided the theoretical framework for the implementation of the newly designed program. Adjustments to the program as implemented, the analysis done in the study, acknowledgement of confounding variables, and ultimate evaluation of the program, including some recommendations, are a direct result of the inherent unpredictable nature of the change process (Haladyna, Nolen, & Haas, 1991).

With a Decatur-generated and administered test at the 9th grade level as a pretest and the first administration of the new 10th grade TAKS Social Studies test as a post-test, Decatur teachers assessed the changes in student achievement over one academic year as they began the implementation of the new model for social studies instruction. Pretesting allowed teachers to identify the students who were most at-risk and to define even more clearly the needs of individual learners. Disaggregation of pretest data led to an in-depth examination of the social studies content strands of the Texas Essential Knowledge and Skills by teachers at all grade levels. One outcome was a streamlined curriculum, which built upon itself in terms of concept and skill attainment. A second outcome was agreement by teachers to be more consistent in their use of social studies vocabulary and specific teaching strategies across grade levels in order to help students build a cohesive picture of the social studies paradigm using familiar thinking
strategies (Decatur Independent School District, 2002-2004). Of particular concern were the low socioeconomic subpopulation and the Hispanic subpopulation (Center on Education Policy, 2003). Early state field tests of the TAKS results showed those subpopulations lagged behind the student population of Texas as a whole (Texas Education Agency, 2001-2002).

Due to my position as curriculum leader for social studies and as director of the Decatur TEKS/Tools Grant Program, collecting systematic data on students and the new professional development and realigned instructional program became a helpful tool for more effective balancing of teacher and students in terms of needs. Teachers developed a clearer understanding of their strengths and areas in need of development. Students’ strengths and needs were likewise identified. Thus, formal program changes in Decatur ISD included focused curricular planning, student intervention, and teacher professional development.

Purpose of the Study

The addition of social studies testing to the menu of high stakes testing that is affecting American high schools found Texas educators unprepared with few models to direct them in planning test preparation curriculum and instruction (Commission on Instructionally Supportive Testing, 2001). Educators are accustomed to testing student achievement in such disciplines as language arts, math, and science (Center on Education Policy, 2002-2003). Social studies presents a much newer and less studied area for the accountability movement.

The relative newness of the TAKS test, its increased rigor, and recent inclusion of the social studies portion of the test as a graduation requirement are variables that make imperative a closer examination of the relationship between professional development of teachers and student scores on the social studies TAKS (Vogler, 2003). After a review of educational trends both
historic and current in the methodology of teaching social studies (Savage, 2003), related unresolved issues regarding effective social studies teaching strategies (Pahl, 2003), and the stakeholder concerns of educators’ failures to deal with the testing barrier (Carter, 2004), the long range implications of high stakes tests and educators’ approach to the curriculum upon which those tests are based indicate that the findings of this study offer useful information for the classroom practitioner. The purpose of this study was to learn whether any statistically significant differences in social studies scores would exist between the treatment schools (those that participated in the TEKS/Tools Training Program) and the non-treatment group schools (those which did not participate in the TEKS/Tools Training Program). The study also assesses the impact of program implementation on the program participant schools’ student performance on the 2005 TAKS Social Studies Test.

Statement of the Problem

Schools receiving the TEKS/Tools Grant Program assumed their changes to the structure of the curriculum and instructional delivery would result in some improvement in student scores on the social studies TAKS test (Social Studies Center, 2001). If that assumption were valid, then there would be an improvement in the social studies TAKS test scores of schools participating in the TEKS/Tools Grant Program as compared to the social studies TAKS test scores of the non-participating schools. There would be an improvement in the Hispanic and low SES subpopulations social studies TAKS test scores in the schools participating in the TEKS/Tools Grant Program as compared to the Hispanic and low SES subpopulations social studies TAKS test scores of the non-participating schools. The program target school (Decatur)
students would show improvement from grades 9 to 10 to 11 on the social studies TAKS test scores.

Method

For this investigation, I developed a quasi-experimental, nonrandomized control group design (Campbell & Stanley, 1963; Glass, 1988; Borg, Borg & Gall, 1998. p.507; Isaac & Michael 1997, p.76). Decatur, the target school, represented a larger group of schools with similar demographic and funding characteristics. Every school in Texas, at the time of the study, had a unique comparison group available for purposes of study and analysis. Since all schools in Texas had the opportunity to be part of the TEKS/Tools Grant Program by virtue of application and compliance with grant requirements, then all schools in that unique comparison group had the opportunity to participate. In this quasi-experimental, nonrandomized study, CCG schools were assigned to the experimental group or the control group based upon whether they participated in the TEKS/Tools Grant Program training and follow-up (Trochim, 2005). Participation identified the treatment; non-participation identified the non-treatment schools. The CCG consisted of the target school, Decatur, and the formula schools from Texas Education Agency (Appendix C) for CCGs. The CCG consisted of the best match under Texas Education Agency criteria (Appendix C); yet, there were individual school differences and limitations. The Decatur target school sample was analyzed in a trend analysis that was appropriate for the 122 subjects in that school using three years of test scores for analysis.

Various statistical analysis techniques were used to analyze the data generated. Gain scores, ANCOVA, and repeated measures analysis were used to examine the overall score gains of the TEKS/Tools Grant participating school groups and TEKS/Tools Grant non-participating
school groups, to examine the scores for statistically significant gains, and to identify effects, interactions, trends or emerging questions for study (Fraenkel & Wallen, 2000).

Research Questions

The first research question: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the Comparison (control) group that did not participate?

The second research question: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between low socioeconomic students in Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the low socioeconomic students in the Comparison (control) group that did not participate?

The third research question: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between Hispanic students in Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the Hispanic students in the Comparison (control) group that did not participate?

The fourth research question: Is there a statistically significant linear trend in the TAKS social study means across the 9th, 10th, and 11th grades for the Decatur students only?
Hypothesis

The Problems as Null Hypotheses

H01: The program group participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade Exit TAKS test scores as compared to the scores of the comparison group that did not participate in the TEKS/Tools Grant Program (control group). (The study compares the mean scores of each school within the treatment and control groups).

H02: The program group low socioeconomic subpopulation participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade Exit TAKS test scores as compared to the scores of the comparison group low socioeconomic subpopulation that did not participate in the TEKS/Tools Grant Program (control group).

H03: The program group Hispanic subpopulation participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade Exit TAKS test scores as compared to the scores of the comparison group Hispanic subpopulation that did not participate in the TEKS/Tools Grant Program (control group).

H04: There will be no statistically significant linear trend in the TAKS social study test means across the 9th, 10th, and 11th grades for the Decatur student’s scores only. (The trend analysis uses the Decatur target school sample individual student scores, n = 122, for 9th, 10th, and 11th grade).

Limitations

Referencing Stephen Isaac and William Michael (1997), I acknowledge the absence of
randomization and the threat to internal validity of variables not reflected in the pretest that might be operating to contaminate the posttest data. Acknowledging that this possibility exists, I made every effort to utilize the formula and CCG format to remediate that possibility (p. 77). Isaac and Michael maintain that validity can be “fairly satisfactory” if groups have similar means and standard deviations, instrumentation is carefully checked, and pre-post score differences are controlled. Regarding external validity, Isaac and Michael see a quasi-experimental design as having the same general questions as a randomized design might have. However, Isaac and Michael posit an advantage to this design, “The non-randomized design has some practical advantages, since it deals with intact classes and does not disrupt the school’s program. By involving a wide variety of classes from several settings, it is possible to achieve an even higher degree of external validity (p. 77).” Borg, Borg, and Gall (1998) address this same issue with the same recommendations, including and encouraging the suggestion of using analysis of covariance to statistically reduce the effects of initial group differences by making compensating adjustments to the posttest means of the groups (p.508). Tabachnick and Fiddell (1996) also support the use of ANCOVA for this purpose (p. 321) in situations where random assignment is not a possibility and there are power concerns.

The study is limited by the lack of an exact match between the Program group and the Comparison group in terms of demographic data and treatment controls. The greater concern is the number of schools that elected to participate in the TEKS/ Tools Program as compared to the number of schools that elected to not participate ($n = 8, n = 29$). The number in the Program group falls slightly under the ideal for quasi-experimental research according to Borg and Gall (p. 229) yet, the remedy of close demographic characteristic similarity “helps a small sample have good statistical power and, thus, can yield important and accurate results (p. 231).”
acknowledge the use of convenience sampling in this study. There exists some attrition and transient student membership in the both the control group and the treatment group. In detailing the subgroups of concern, I acknowledge the reality of underreporting the number of low SES students based solely on the criterion of application for and participation in the free or reduced student lunch program. In addition and with some emphasis, I must acknowledge certain bias resulting from decades of practice in the classroom teaching social studies as well as involvement in the grant process for the Decatur target school.

Definitions of Terms Used in Study

1. Academic Excellence Indicator System (AEIS): AEIS consists of a wide range of information on the performance of students in each school and district in Texas that is available each year reflecting the previous year’s performance indicators. The indicators include: TAAS passing rate by grade (to be replaced by TAKS in year 2004), end-of-course exam information, attendance information, dropout rates, percent of students in advanced courses, advanced placement and international baccalaureate results, and SAT /ACT results. Performance on each of those indicators is shown disaggregated by ethnicity, special education, and income status. AEIS reports also provide extensive information on school district staff, finances, programs, and demographics.

   2. Campus comparison group (CCG): CCG is a “unique” comparison group of up to 40 campuses identified as a match for an individual school. CCG is constructed using the most dominant characteristics of the target campus. In the research study, the target campus is Decatur High School. Initially, the most similar 100 campuses were selected, and then each campus was refined by selected characteristics. Six characteristics predominate in the analysis.
They include: percentage African American, percentage Hispanic, percentage White, percentage economically disadvantaged, percentage LEP (limited English proficiency), and percentage mobile. Other considerations are used in the construction of the cohort groups. The CCG was constructed for any target school as a vehicle to compare any number of accountability factors. They were constructed by TEA using a set formula. Individual schools can then refine the CCG further for a variety of purposes. CCG allows a school to compare itself to similar schools rather than use the state population as a whole or use other disparate schools for comparison (see Appendix C).

3. Comparison group: The name referring to the schools in the CCG that elected non-participation in the TEKS/Tools Grant Program. Using a quasi-experimental model, this group of schools becomes the control group. After verification and deletion due to erroneous reporting and other issues as detailed in chapter 3, this group numbered 29 schools.

4. Decatur sample: The name for the Decatur target school as used for trend analysis with the 9th, 10th, and 11th grade scores of 122 students.

5. Education service centers (ESC): Twenty education service centers were established in 1967 by the Texas Legislature to provide information, services, staff development and other support to the schools of Texas. ESCs serve as field structures for TEA in the furthering of special programs, grants, training and follow-up and other duties. Assignment of specific projects by TEA can be found on the web and these are adjusted annually as needed.

6. English language learner (ELL): The name given to any student whose first language is not English. Texas still uses the term English as a second language or ESL; however, ELL is the term used by most national entities including NCLB.
7. End-of-course (EOC) test: The social studies end-of-course tests were used from 1997 to 2002 to assess minimal skills in the social studies disciplines for secondary students. The tests had little impact on school accountability and were used by Texas Education Agency and local districts for study and analysis. The domain areas and other information from the tests were used in the formulation of the new Texas Assessment of Knowledge and Skills (TAKS) social studies tests.

8. Exit testing: This term refers to the exit exam for high school graduation. In Texas, it is the 11th grade TAKS exam. The term is used interchangeably with the term “high stakes testing” due to the diploma-dependent nature of the exams.

9. No Child Left Behind (NCLB): This term refers to the No Child Left Behind Act of 2001 (US Department of Education, 2001). The new law reflects a consensus-first articulated in President Bush’s January 2001 No Child Left Behind framework -on how to improve the performance of America’s elementary and secondary schools while at the same time ensuring that no child is trapped in a failing school. The legislation reauthorizes the former framework, the Elementary and Secondary Education Act of 1965. The NCLB stresses increased accountability, more choice for students and parents, greater flexibility for local entities, and places primary importance on reading.

10. Program group: This refers to the CCG schools that elected to participate in the TEKS/ Tools Grant Program. The participants had three separate verifications: by TEA for grant monies dissemination, by the Social Studies Center for training and follow-up purposes, and by myself in interview format. Using a quasi-experimental model, this group of schools becomes the experimental group. After verification and deletion due to erroneous reporting and other issues as detailed in chapter 3, this group numbered 8 schools.
11. Social Studies Center (SSC): The Social Studies Center was a temporary agency within Texas Education Agency, formed in 2000 with special funding for the purpose of initiating specific programs and grants in preparation for the TAKS tests in the core area of social studies. Funding for the center ceased in 2004 and the skeletal SSC was folded into the TEA offices in Austin. A central function of the SSC was the TEKS/Tools Grant Program, the funding of the grants, and the training and monitoring of the participating schools.

12. Socioeconomic status (SES): This term references the economic status of the student’s family and generally is used to reference students of concern in a low SES. A common indicator of low SES is a student’s qualification for free or reduced price lunch. I must acknowledge that students may be low SES and have families that do not apply for the free or reduced lunch program. TEA also refers to the category as economically disadvantaged in many instances.

13. Texas Assessment of Academic Skills (TAAS): The TAAS measured the statewide curriculum in reading, mathematics, and writing at the exit level for students who were enrolled in grade 9 or higher on January 1, 2001. It has been supplanted with the TAKS test.

14. Texas Assessment of Knowledge and Skills (TAKS): As mandated by the 76th Texas legislature in 1999, the TAKS would be administered beginning in the 2002-2003 school year. The TAKS measures the statewide curriculum in reading at grades 3-9; in writing at grades 4 and 7; in English language arts at grade 10 and 11; in mathematics at grades 3-11; in science at grades 5, 10, and 11; and social studies at grades 8, 10, and 11. The Spanish TAKS is administered at grades 3 through 6 only so testing in Spanish is not applicable at the high school level. Satisfactory performance on the TAKS at grade 11 is a prerequisite for a high school diploma.
15. Texas Education Agency (TEA): is the administrative unit in Texas for primary and secondary education. The agency is comprised of the commissioner of education and agency staff with duties that include guiding and monitoring activities and programs related to public education. The stated mission of the agency includes the goal of providing all students with a quality education.

16. Texas Essential Knowledge and Skills (TEKS): The TEKS is the required State of Texas curriculum as stipulated in Texas Education Code, chapter 28. It consists of foundation and enrichment subjects. The foundation subjects include language arts and reading, mathematics, science, and social studies. All other subjects are enrichment. Districts are required to provide instruction in the essential knowledge and skills of the appropriate grade levels in the foundation curriculum. Districts are to use the essential knowledge and skills in the enrichment curriculum as guidelines for instruction.

17. TEKS/Tools grant program- The minimum two-year training and implementation plan for districts under TEA guidelines that stressed TEKS-TAKS alignment in social studies and use of prescribed instructional strategies to deliver that aligned curriculum. The competitive grant was monitored through a temporary agency of TEA, the Social Studies Center, and training was done through the ESC system in Texas (see Appendix B). For purposes of this study, the TEKS/Tools Grant Program participation is the intervention treatment in the quasi-experimental model. Participation in the program determined the assignment of a school to program or comparison groups.

18. Texas learning index (TLI): The TLI is a statistic that allows for comparison both across years and across grades within a subject area for reading and mathematics. It can be used
to assess learning progress across test administrations. The index is an integral part of the formula used for constructing the CCGs.

Assumptions

I assumed that the 10th grade TAKS social studies test and the 11th grade TAKS social studies test for the class of 2005 are reliable and valid measures of the state curriculum (Appendix D). Note: I was part of the state test development process as a teacher member of the development committee (See Figure 1 Research and Study Timeline).

I assumed, and verified to the best of my ability, that the program changes within the Decatur ISD and other treatment group districts included professional development for all social studies teachers and curriculum work, such as alignment to the TEKS. I assumed that the target school, Decatur, proceeded in the manner that all compliant schools were assumed to have proceeded to address their grant requirements as a stipulation for funding. The assumption is based on the school districts’ verification of completion of the TEKS/Tools Grant requirements for grant funding. I further verified program changes and adherence to the terms and conditions of the TEKS/Tools Grant by confirmation from state supervisors and interviews with the districts participating and assumes that information to be accurate.

I assumed that the research design sample school pretest given at 9th grade (see validity information under limitations chapter 1 and in chapter 4), 10th grade TAKS social studies test, and the 11th grade TAKS social studies test are aligned and increase in degree of difficulty appropriate for the level being tested (Appendix D). I assumed that a student’s test awareness regarding test format and ability to test increase with each sample test taken.
I assumed that the CCG of 40 schools was the best possible match for the Decatur target school as determined by formula and parameter, within the public high schools in Texas (the process used for defining the CCG is detailed in Appendix C.)
CHAPTER 2

REVIEW OF LITERATURE

The Case for Standards

Why are standards important? There appear to be three principle reasons advanced for the development of standards: standards serve to clarify teaching goals, raise expectations, and standards provide a common set of expectations (Ravitch, 1995). Aside from funding, is there a more hotly debated issue than educational standards and assessment (Berliner & Biddle, 1995)? In the debate, standards and assessment seem inseparable. Notably many schools and the stakeholders interested in their success use the terms standards, goals, and objectives somewhat interchangeably. Thus, the term “standards” is very context specific.

From the pandemonium of voices debating the nuances of standards and assessment, former Assistant Secretary of Education Diane Ravitch is commonly recognized as one of the chief architects of the modern standards movement with a moderating view of the subject. In her book National Standards in American Education: A Citizen’s Guide (1995), Ravitch provides a common-sense rationale for standards:

Americans . . . expect strict standards to govern construction of buildings, bridges, highways, and tunnels; shoddy work would put lives at risk. They expect stringent standards to protect their drinking water, the food they eat, and the air they breathe. Standards are created because they improve the activity of life (pp. 8-9).

Ravitch (1995) asserts that just as standards improve the daily lives of Americans, so, too, will they improve the effectiveness of American education. She posits, “Standards can improve achievement by clearly defining what is to be taught and what kind of performance is expected" (p. 25).

Many people, according to public surveys, apparently share such a view. The polling firm Public Agenda conducted a number of surveys on the issue of standards over the last several
years. They found that most Americans strongly support higher standards that are clear and specific (Farkas, Friedman, Boese & Shaw, 1994), believing that higher expectations produce better performance. Teachers, as well, support proposals to raise standards, which they expect to improve their students' academic performance (Johnson & Farkas, 1996). Findings indicate that students also see value in standards, saying that higher standards will make them work harder, and they expect to learn more as a result (Friedman & Duffet, 1997).

Standards and assessment have been inextricably linked by proponents of standardized assessment to rally support for public policies that use high-stakes tests to change the behavior of teachers and students in desirable ways to meet those standards (Grant, 2001b). But the use of gatekeeper tests is not new, and their effects are not always desirable (Burroughs, 2002). Stakes, or the consequences associated with test results, have long been a part of the American scene. For example, early in the 20th century, scores on the recently formulated standardized tests could, for immigrants, result in entrance to or rejection from the United States of America. In the public schools, test scores could uncover talent, providing entrance into programs for the gifted, or as easily, provide evidence of deficiencies, leading to placement in vocational tracks or other alternative placements. Test scores could also mean the difference between acceptance into or rejection from the military. And throughout early twentieth century society, standardized tests scores were used to “confirm” the superiority or inferiority of various races, ethnic groups, and social classes. Used in that way, the consequences of standardized tests insured maintenance of the status quo along those racial, ethnic, class lines. So, for about a century, significant consequences have been attached to scores on standardized tests (Brookover, 1979).
A Recent History of High-Stakes Testing

Our current state of faith in and reliance on tests has roots in the launch of Sputnik in 1957. Our (then) economic and political rival, the Soviet Union, beat the United States to space, causing our journalists, politicians, educators, and the general public to question American education with additional vigor. At that time, state and federal politicians became more actively engaged in the conduct of education, including advocacy for the increased use of tests to assess school learning (Kreitzer, Madaus, & Haney, 1989).

The publication in 1983 of the controversial report *A Nation at Risk* (US Department of Education, 1983) signaled a warning cry to educators, parents and legislators about the state of education in the United States. The report, which purported to demonstrate conclusively that children in the United States were far behind their peers in other countries, sparked a firestorm of education reform that continues today. A major offshoot of that reform was the reauthorization of the Elementary and Secondary Education Act as the Improving American Schools Act (IASA). IASA included requirements for states to establish high standards and to create and administer standards-based assessments for “all students.” Those assessments often are called “high-stakes” assessments, in part, because many states require a passing score in order for students to graduate from high school. The Goals 2000 Educate America Act (Goals 2000) and IASA were among Congress’ major responses to *A Nation at Risk*. Together, the acts stand for the principle that all children can learn and achieve high standards and are entitled to participate in a broad and challenging curriculum.

The acts also establish requirements for states accepting their funds. Namely, states must create and implement education improvement plans that include processes for developing and adopting curriculum standards for “all students.” To help measure progress, Goals 2000 requires
that state improvement plans include a process for developing and implementing reliable state educational assessments. The assessments must be aligned with state curriculum standards and involve multiple measures of student performance. Moreover, the results of the assessments must provide interpretive and descriptive information and must be disaggregated within each state, local school district and school to provide a comparison of scores for subgroups of students. Congress intended the disaggregation to “help schools ensure that all types of students are making progress towards meeting the state standards” (Goals 2000).

The current emphasis on student assessments, to measure the quality of education that students receive in each school, district, and state, had its genesis in September 1989, when President George H. W. Bush joined the nation's 50 governors to convene an education summit at Charlottesville, VA. A National Governors Association (2003) policy states, "Governors understand the importance of collecting and using student data to guide decision making. As a result, Governors support the creation of a semi-autonomous group in developing appropriate, credible, and reliable indicators to measure performance" (p.10).

A National Education Goals Panel was established to issue annual reports on progress of the nation and states toward the goals. Six resource groups of national experts suggested specific data that would objectively measure progress toward the goals. One of those recommendations was to measure progress against voluntary national education standards. Governors held eight regional public hearings across the country in spring 1991 to hear comments from educators and the public on the selection of indicators.

By 1993, a goals panel described the need for a national student data reporting system for assessing students' completion of school. Another panel recommended criteria and procedures that might be used to certify national education standards. In March 1994, President Bill Clinton
signed the Goals 2000 Educate America Act, codifying the eight national education goals and giving the goals panel new responsibilities.

In 1996, the Goals Panel's annual report focused on two areas of education reform of great interest to states and local communities: standards and assessments. The panel stated that it would be impossible to achieve the goals unless states and local communities demanded more student achievement and designed new forms of assessment to determine whether students have mastered challenging subject matter. The recently reenacted Elementary and Secondary Education Act, known as No Child Left Behind, is a further decisive step in creating a process for national standards of educational achievement.

Three days after taking office in January 2001 as the 43rd President of the United States, George W. Bush announced No Child Left Behind, his framework for bipartisan education reform that he described as “the cornerstone of my Administration” (US Department of Education, 2004). President Bush emphasized his deep belief in our public schools, but also a great concern that “too many of our neediest children are being left behind,” (p.1) despite the nearly $200 billion in federal spending since the passage of the Elementary and Secondary Education Act of 1965 (ESEA). The President called for bipartisan solutions based on accountability, choice, and flexibility in federal education programs. Less than a year later, President Bush secured passage of the landmark No Child Left Behind Act of 2001 (NCLB Act). The new law first articulated in the President's No Child Left Behind framework speech in January of 2001, emphasized improvement of the performance of America's elementary and secondary schools while at the same time ensuring that no child would be trapped in a failing school.
The NCLB Act, which reauthorizes the ESEA, incorporates the principles and strategies which included increased accountability for states, school districts, and schools; greater choice for parents and students, particularly those attending low-performing schools; more flexibility for states and local educational agencies (LEAs) in the use of federal education dollars; and a stronger emphasis on reading, especially for our youngest children.

Although we can safely say that social studies TAKS testing in Texas is not a specific requirement of the NCLB Act, we must acknowledge the inherent implications of the standards and testing movement. So we turn our attention from standards and testing in generalities (or the specifics of math or science as referenced in Nation at Risk and Goals 2000) to that very arena of concern: high stakes social studies testing.

Focus on Social Studies Standards and Assessment

In the post-9/11 world, it is more important than ever for young Americans to learn the history of their nation, the principles on which it was founded, the workings of its government, the origins of its freedoms, and how the U. S. has responded to past threats from abroad. What educator in the United States would disagree with this premise? Educators seem to agree that a well-crafted K-12 curriculum must assure that students are deeply immersed in U.S. history (as well as civics, geography, world history, and other social sciences) and that high school graduates be knowledgeable about America's past (Elazar, 1984). Though schools cannot be held exclusively responsible for forging good citizens—that solemn duty is shared by parents, places of worship, and myriad other institutions—they have a unique obligation to handle the "cognitive" side; i.e., to make certain that young people gain the requisite knowledge and intellectual skills.
In this era of standards-based reform, the subjects most apt to be taken seriously and taught well in our schools are those for which the state sets high quality standards that make clear what teachers are expected to teach and children to learn (Firestone, Camilli, Yorecko, Monfils, & Mayrowetz, 2000); those that are tested, and tested with rigor (Grant, Derme, Gradwell, Lauricella, Pullano, & Tzetzo, 2001), and those for which the accountability system confers rewards or sanctions on students, educators, and schools according to how well they have succeeded in teaching and learning (Grissmer, Flanagan, Kawata, & Williamson, 2000).

In that context, however, U.S. history has not fared well (Grant, 1997a). Although almost every state requires students to enroll in at least one course in the subject (typically in eleventh grade), history has seldom, if ever, appeared in any statewide testing and accountability systems (Eilperin, 2001). Yet, assessment after assessment and study after study show that history is the core subject about which young Americans know least. The fraction of students (in grades 4, 8 and 12) who reach the "proficient" level on tests administered by the National Assessment of Educational Progress (NAEP) is smaller in history than in any other area or field (National Center for Education Statistics, 2004). Of the 25 states that have or intend to have high school exit exams by 2008, only nine include any area of social studies among the subjects tested (Gaylor, 2003) See Appendix A for detail. At the epicenter of the continuing and often acrimonious debate about what our children should learn in school has been the belief that rigorous state standards for history and social studies could significantly enhance both teacher preparation and student achievement (Leming, 2003).

Testing is a reality of public school teaching for which teachers must be prepared. Because the social studies arena is relatively low in the testing hierarchy, it is even more in need of paradigms for effective teaching and management. According to The Center on Education
Policy (2005) High School Exit Exams: Report, innovative programs and policies are just now beginning to spring up in states with exit exams. The report also notes vast differences and inconsistencies in implementing programs when instructional models are new to the field and have not been tested, much less, proven. States face many challenges as they design programs and strategies that will effectively scaffold learners and improve their knowledge bases in social studies. Typically, areas with the greatest accountability are the most researched and resourced. At this time, the number of states testing social studies as compared to the other areas tested shows less weight and importance in priority. Only 9 states in 2005 had diploma-dependent social studies tests compared to 19 with math and reading tests. Future testing does not improve the outlook. By 2009 the projection out of 25 states testing at the high stakes level is: language in 25 states, math in 25 states, science in 14 states and social studies increasing to 11 states (Center, 2005).

Today, the federal No Child Left Behind (NCLB) Act of 2001 is the strongest force driving United States schools toward standards-based reform and stronger pupil achievement. Yet, a 2004 Kappan poll indicates respondents believe: they are largely uninformed regarding NCLB; achievement should not be measured on the basis of single area tests; and, that math and English testing alone is not a dependable indicator for improvement (Kappan, 2004, tables 6-11). Without intending to, NCLB may actually worsen the plight of U.S. history as an academic subject. By concentrating single-mindedly on reading, math, and science, it will likely reduce the priority that states, districts, and schools assign to other subjects. And by highlighting performance (or the absence thereof) in only those three core fields, the attention of state and community leaders will likely focus on their schools' results in those subjects-and deflect their attention from others. Manzo (2005) argues, “The unintended consequence of No Child Left
Behind has been to put history into an even more marginal position” (p.1). A problem, yes, but one that states and schools can solve if they choose to do so. Maryland state schools superintendent, for instance, has appointed a task force to study the “state of social studies” education statewide to determine recommendations to strengthen the teaching of the subject. If state tests become more universally high stakes for social studies, states will be forced to make improvements, and merely calling social studies a core subject will not be enough to improve instruction in many educators’ views (Firestone, Mayorwetz, & Fairman, 1998). Social studies educators have long believed the social studies disciplines should receive equal emphasis with other core subjects. Moreover, forty-eight states (all but Iowa and Rhode Island) and the District of Columbia have previously established academic standards in social studies, signifying they have at least gone through the motions of detailing what they expect their teachers to teach and students to learn in the field (Saxe, 1997).

The social studies discipline has a unique place on the menu of high stakes tested areas. The standards for testing remain controversial. Contemporary observers (Crocco, 2004; Libresco, 2004; Segall, 2002; S.G.Grant, 2001) seem in agreement on the controversial and conflicted nature of testing an area that is charged politically. The debate and discussion revolve around topics, individuals to be studied, and subject area focus for testing as well as method and format for testing.

As part of the on-going culture war in the U.S. in 1996, conservatives led by Lynn Cheney and her supporters challenged the specifics of the history standards as developed by the National Center for History in the Schools at the University of California under the guidance of the National Council for History Standards ("History Standards," Education Daily, January 1995). The standards were developed with funding from the National Endowment for the
Humanities and the U.S. Department of Education. The History Standards Project, directed by the National Center for History in the Schools (NCHS), first published three sets of standards: *National Standards for History for Grades K-4, National Standards for United States History,* and *National Standards for World History* (NCHS, 1995). Publication of the standards drew immediate criticism from Lynn Cheney who, as former head of the National Endowment for the Humanities, had approved funding for the project. Others joined the debate, either condemning the history standards outright or making recommendations for their improvement. Of the mainstream tested areas: language arts, math, science, and social studies, all have controversies inherent. However, social studies remains politically volatile as evidenced by the debate and discussion included in the 2004 American Educational Research Association annual meeting sessions relating to history teachers and teaching, particularly the sessions dealing with trends and issues. (AERA, 2004).

An examination of the testing situation from state to state reveals that Texas and New York constitute the apex of testing rigor and expectation (Pahl, 2003). The inclusion of social studies in states’ high stakes testing has been accompanied by a rigorous upgrade for social studies state curricula overall (Gaylor, Chudowsky, Kober, and Hamilton, 2003). This is not to say that with No Child Left Behind, that classroom or district emphasis will be as rigorous as the newer state curricula might suggest. Texas added social studies as a tested area early (relative to the other states) resulting in serious curricular implications (Haney, 2001). As Appendix A indicates, Texas developed a rigorous test; yet, teachers and students had little time to prepare for the test compared to their preparation time for other subject area tests.
History of Accountability Testing in Texas

To better understand the development of the social studies TAKS test in Texas, a review of the history of accountability testing in Texas is in order. Specifically, Texas has had a series of testing experiments dating back to 1979. Educators point to the Texas Assessment of Basic Skills Tests as the beginning of high stakes testing in Texas because the test results were reported and formed the beginning of accountability for Texas districts. A discussion timeline of testing in Texas follows (Texas Education Agency, 2004).

In 1979, TABS or the Texas Assessment of Basic Skills Test was formulated. The legislature passed a bill requiring basic skills competencies in math, reading, and writing for grades 3, 5 and 9. Because there was no state-mandated curriculum at that time, the learning objectives for the TABS were created by committees of Texas educators. In 1983, the Texas legislature began requiring retesting. Although TABS was not a “diploma-denial test,” 9th grade students who did not pass the test were required to retake the exam each year thereafter while in school. Because results were reported, the TABS test was viewed as the beginning of high stakes accountability for school districts. Social studies testing at that time was not a looming concern. Consequences were minimal but accountability became a topic of discussion statewide.

The next major step in the developing accountability forum was the 1984 TEAMS or Texas Educational Assessment of Minimum Skills Test. Thus, the class of ‘87 became the first class in which students were required to pass the exit level exam in order to receive a diploma. A 1983 Texas legislature resolution changed the wording of the Texas Education Code, requiring the assessment program to measure “minimum skills” rather than “basic skills competencies.” The TEAMS test began in the 1985-86 school year, replacing the TABS. The TEAMS increased the rigor of the state assessment and added individual student sanctions for performance at the
exit level. TEAMS tested math, reading, and writing, and was administered to students in grades 1, 3, 5, 7, 9 and 11, with the 11th grade testing being the “exit level” assessment. Again, social studies was not an area of focus in any TEAMS tested venue.

The next step in the progression of testing accountability was the 1990 Texas Assessment of Academic Skills Test, which was a criterion-referenced test. The TAAS test shifted the focus from minimum skills to academic skills, which represented a more comprehensive assessment of the state-mandated curriculum, the Texas Essential Elements. TAAS assessed higher-order thinking skills and problem-solving in math, reading and writing for grades 3, 5, 7, 9, and 11 exit level. The State Board of Education considered the following factors when establishing the levels of satisfactory performance. First, the TAAS assessed a broader range of the Texas essential elements than TEAMS did. Second, in comparison to TEAMS, the TAAS test items were more difficult. Third, the TAAS served multiple purposes by providing scores and consequences at the student level, the school level, and the district level. Due to those factors, the board set a one-year interim standard for satisfactory performance. From 1992 to 1993 TAAS transitioned from a fall to a spring testing program, and in 1993-1994 assessment was expanded to include grades 3-8 in reading and math. The writing test was moved to grades 4 and 8, and the exit level test was moved from grade 11 to grade 10.

In 1993, the legislature enacted the creation of a new statewide integrated accountability system that included the rating of campuses and districts. The inclusion of TAAS in the accountability system, the public release of performance results, and the exit-level requirement for graduation made TAAS the most high stakes assessment in Texas’ history up to that year.
In 1994 the board voted to align the passing goals at grades 3-8, with the goals being established at the exit level. The new standard, the Texas Learning Index (TLI), allowed comparisons of achievement across grades while maintaining the same passing goals for exit level students. The TLI helped districts determine whether each student was making the yearly progress necessary to meet minimum expectations on the exit level reading and math test in 10th grade.

In 1995, for the first time, an accountability measure was linked to a social studies assessment starting with the eighth grade TAAS. Development of end of course science and social studies tests for high school began during the 1995-1996 school year, were field tested in the spring of 1997 and were administered in spring of 1998 as “benchmark administrations.” The results were viewed as “voluntary adequacy of preparation” results and used by local districts and TEA for study and for future assistance in TAKS tests preparation.

1999 TAKS (Texas Assessment of Knowledge and Skills) Test

In 1999, the introduction of the Texas Assessment of Knowledge and Skills (TAKS) test began, and with it the development of a defining social studies standards-based test. The legislature passed bills ending social promotion and creating a more rigorous testing program (Texas Education Code, Chapter 39 and 28 respectively). As mandated by the 76th Texas Legislature, the Texas Education Agency began to develop a new assessment program, the Texas Assessment of Knowledge and Skills (TAKS) that was to be aligned with the state-mandated curriculum, the Texas Essential Knowledge and Skills (TEA, 1999).

Under the new law, students in grades 3 (reading), 5 and 8 (reading and math) were, and still are, required to demonstrate proficiency on a state assessment test, and achieve passing
grades in order to advance to the next grade level. At the 11th grade (reading, writing, math, science and social studies) students must pass the TAKS test, in addition to receiving the required number of credits, in order to receive their high school diploma. The Texas Education Code (TEC) charges the State Board of Education with establishing the passing standards (performance standards) on the new TAKS test (Camilli, 2000). Social studies testing has now reached the basic skills or minimum competency level and has evolved into a higher level thinking criterion-based test (Center on Education Policy, 2000-2002).

2002 through 2005 TAKS Phase-in

In the spring semester of 2002, the last administration of the TAAS test occurred. (Although Exit level students who failed any subject area test would continue to retest on TAAS.) TAKS was field-tested across the state of Texas and became the new statewide assessment program administered beginning in the 2003 school year. From February to May, 2002, statewide field-testing for grades 3-11 was conducted in order to collect student performance data on test items. In November of 2002, the State Board of Education set passing standards for the new TAKS test. In the spring of 2003, the first live administration of the TAKS test, which generated scores that counted for students, occurred. The exit level or 11th grade TAKS scores became diploma dependant in the spring of 2004. Currently Texas is one of only nine states that mandate a social studies test as an inherent part of a diploma-dependent battery of tests (Center on Education Policy, 2003). Although Texas has been involved in standardized testing and accountability issues for several decades, recently Texas schools have been viewed as models for equity, progress and accountability (Hurwitz, 2000). Texas teachers feel pressure and
any study or enlightenment on this topic will hopefully assist in goal attainment and positive practical application of ideas to enhance student achievement.

**Issues and Concerns Regarding High Stakes Testing**

Curricular issues abound regarding high stakes testing. The states that currently require high school graduation or exit exams include: Alabama, Alaska, Florida, Georgia, Indiana, Louisiana, Maryland, Massachusetts, Minnesota, Missouri, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, South Carolina, Tennessee, Texas, and Virginia. Arizona, California, Idaho, Utah, and Washington are phase-in states and will have diploma dependent tests by 2008. Relatively early reports, such as the Report of the National Commission on Testing and Public Policy (1990), opined that test scores should not be used as a sole means for making decisions about individuals or institutions, since most tests are imperfect measures of performance. They recommended using a menu of tests to inform instruction, evaluate programs, and determine school accountability. Those findings are contrary to the use of largely singular high stakes test scores as an indication of both individual and school accountability, as well as, program evaluation. As the late U. S. Senator Paul D. Wellstone said, “High stakes tests are a harsh agenda for America’s children” (High Stakes, 2000).

There is little real research to suggest that adding rigorous tests to secondary education has, as yet, had any motivating impact on a student (Carnoy & Smith, 2000). In fact, studies indicate the potential for just the opposite for the performance of at-risk groups. Failure that is attributed to a student’s lack of ability is extremely hard for a student to overcome (Alderman, 1990). Yet, little is built into the preparation for testing that increases a student’s efficacy or allows for differences in learning or allows for error to be a natural part of the learning cycle due
to time and other pressures (Amrein & Berliner, 2002). Teachers believe that most of their students can reach mastery learning if adequate time and appropriate instruction are provided (Swope, 2000). Yet, few teachers believe that criterion is met with the emphasis on high stakes testing, as it currently exists (Stipek, 2005).

The New York Regents Tests best parallel the Texas tests and studies analyzing scores from the tests reiterate the concern for various at-risk groups (McNeil & Valenzuela, 2001, McNeil, 2000). Achievement differences amongst students are a reality for all educators (Madaus & Clarke, 2001). Data collected in 2004 show that fewer than a third of Hispanic students in New York earned a high school diploma in four years, the worst showing of any state, compared with more than half of all Hispanic students in the nation as a whole (Winter, 2004). Other states replicate that trend (Jones, Hardin, Chapman, Yarbrough, & Davis, 1999). Half of the minorities (excluding Asians) failed the Utah Basic Skills Competency Tests, but only 1 in 5 Anglo or Asian students failed on the tests (Utah Legislature Interim Committee on Education, May 2005). Texas Education Agency announced that 9% of class of 2005 did not graduate with their peers in Texas due to failing one or more parts of TAKS (TEA, May 2005). (The ethnic and socioeconomic makeup of that 9% is not available for analysis at the present time; however, the figure constitutes an improvement from the projected third that TEA expected to fail one or more of the tests.) Notably those students have had several chances now to focus on the tests or test that they did not pass in their retake performance. The original projection did not predict the retake success rates. Educators have demonstrated concerns as to the ethnic and socioeconomic makeup of that 9% who did not make the grade (Orfield & Kornhaber, 2001). Overall Texas has a larger disadvantaged population than the nation with 46.2% low SES as compared to 38.5% nationally (Standard and Poors, 2005). Texas Education Agency figures
vary somewhat according to data source. Effects of high stakes tests on dropout rates have been covered in several works that indicate that low SES students have a significantly higher chance of dropping out (regardless of achievement) in states with high stakes tests for graduation (Kossan, 2000; Amrein, 2002). Using data from National Educational Longitudinal Study, Sean Reardon and Claudia Galindo at the Population Research Institute at Pennsylvania State University reported at the 2002 AERA annual meeting in New Orleans evidence that high stakes tests do affect students’ decisions to drop out of school. Another report prepared for the New York City Department of Education verifies similar findings with a four-year longitudinal study of rising dropout rates (Rankin, 2003). Data reports from the New York State Education Department echo that unfortunate trend (2003-2004).

A 2005 report from researchers at Stanford University suggests that using a multiple measures approach would provide a more sound evaluation of student progress for graduation than the single high stakes test that states presently use. Stanford University educators led by Linda Darling-Hammond (2005) expressed several concerns in their report on high stakes single-test graduation gate-keeping: reduced graduation rates (especially for ethnic minorities and other at-risk students), incentives for schools to push out students who do poorly when school ratings are contingent on the average pass rates of students, and a narrowing of the curriculum including neglect of higher order performance skills where limited test measures are used. Darling-Hammond states, “High school graduation policies have important consequences for teaching, learning, and student achievement” (p. 38). As lead author of the report, Darling-Hammond adds, “It is important both to balance tests with other sources of evidence and to encourage students to do real-world tasks that go beyond what can be measured with multiple-choice questions” (p. 42). Ultimately, educators will continue to ask the questions posed by Neill and
Gaylor (2001) in their thought provoking overview, “Do high stakes graduation tests improve learning outcomes? Are we raising standards or raising barriers?”

Social studies high stakes testing inherently manifests many of the same concerns that other disciplines do (Muller & Schiller, 2000). Due to the nature of the content taught in social studies, curricular and instructional decisions will vary by worldview and will always be subject to interpretation (Segall, 1999). Specific concerns have arisen regarding how to help teachers best prepare the students for a high stakes social studies test (Segall, 2004). Avner Segall has stressed the need to re-think theory and practice in the classroom and the role of all aspects of teacher preparation. The suggestion of refocusing on the instructional methods cannot be overstated when one reviews Segall’s work. In Social Studies Education: A Re-conceptualized Framework (2002), Segall argues, we have not come very far in our approach from the traditional social studies pedagogy of fifty years ago. The relevance, big ideas, connected and interdisciplinary approaches, the language used, the relationships and the complexities cited as effective are still not employed in the classroom. And, the high stakes nature of the courses now dictates that educators must infuse meaning, interest and consistency into the curriculum. Other contemporary observers call for a “community of inquiry” into the study of big ideas (Crocco, 2001). S. G. Grant (2000) examines the gap between national and state reform goals and the attention to how teachers should prepare their students. His study focuses on the New York Compact for Learning, a statement of what resources, conditions, and practices are necessary for all students to learn so that they will have success on the New York Regents social studies tests. He concludes, “None of these efforts, however, seriously addresses how experienced teachers will learn the intended innovations” (p.1). Grant indicates a need to study and assess what works to prepare teachers for their goal of educating their students.
Lack of Research and Quality of Research

A recurring theme in the review of literature regarding high stakes testing and, particularly, the area of social studies testing, is the lack of research. There is also some very vocal criticism of the quality of the research available. No data in the NCES data bank regarding social studies score results by state are available as yet (National Center for Education Statistics, 2004-2005). NAEP has data on history testing done prior to 2001; there has been no analysis since and none of the high stakes exit tests have been analyzed (National Assessment of Educational Progress Project, 2004-2005). The National Assessment of Educational Progress (NAEP) reports scores for 12th graders as unchanged from the 1994 testing of history (2004). A complete demographic breakdown of testing high school students in history (covering ethnic and low SES) was last compiled in 1992 (NCES, 2003). The NAEP reporting methods and emphasis has been questioned (Johnson, 1992). Currently, there is much debate over the quality of education research. Critics’ responses to it continue to be controversial; Education Week and other periodicals have labeled it inadequate in terms of data collection and peer review and criticized the reporting results (Viadero, 2005). Yet, the need for more study in the area of high stakes testing and, especially, social studies testing is critical. Most practitioners want answers now. The delay of data collection by states and other entities and the analysis of that data are problematic for the practitioner looking for guidance.

Other areas of concern regarding research on high stakes tests are: How are tests scores evaluated? How are they analyzed? What statistical procedures are used and what meaning is given to the findings? Are scores used for the purpose of rating schools and districts? We know we evaluate students but are scores also used for evaluating teachers? How are variables that affect scores treated in the equations? Are there consistent ways to utilize those scores? Who
determines the use or abuse of playing the score game for schools (Grant, 2001)? Inherent problems include corruptibility or inflation of test scores due to teacher emphasis on the test format rather than an appropriate curriculum, corruptibility of the process of interpretation of the scores such as actual manipulation of the data, and contrived attempts to make causal links between the scores and various educational programs (Heubert & Hauser, 1999). The skimpy literature on those topics leaves educators with many more questions than answers. Score evaluation is a dilemma with which each testing entity deals differently. New York, the prototype state for high stakes testing, has recently altered the manner of score reporting for school rating purposes. New York City Schools Chancellor Joel Klein announced a modified plan in June 2005 (Herszenhorn, 2005). Klein calls the method a “sophisticated analysis known as value added” (Herszenhorn, 2005). In fall 2005, New York City will begin using the value-added approach to analyze standardized test scores to determine schools’ effectiveness in helping students. Officials will compare the change in student scores from year to year in one school against the performance of students in others to get a clearer picture of how students are progressing. The procedure appears to be similar to the TLI that Texas used for TAAS evaluation, which takes into consideration the gains a student makes within a given year and relates that to grade level expectations. The stakes are high for students and for teachers, administrators and all school stakeholders as well (Domench, 2000; Durbin, 2000). Houston Independent School District officials announced in June 2005 that they are considering bonuses for teachers in low performing schools whose students post measured improvement gains (HISD, 2005). The bonus system has been discussed by districts around the country and is fraught with numerous concerns (Administrative bonuses, 2001). In early November 2005, Texas Governor Rick Perry announced a plan for teacher pay raises linked to improvement in 100 low achieving
schools. Reaction was predictably mixed with concern as to methods of evaluation and fairness issues.

Controversy and conflicts are very much part of education today and the stakes are high for all involved. Educators need more information to make the difficult curricular and instructional decisions to prepare students for high stakes tests. Fall 2005 saw the emergence of a “new” approach in the establishment of the Southern Methodist University School of Education and Human Development. This school, according to its stated mission (2005) was established to “teach programs that are aligned with President Bush’s hallmark education program NCLB to address the achievement gap” (SMU, 2005) To identify teaching techniques that might lift achievement, SMU is using federal and private money to conduct research in the local system that mirrors the kind of experiments conducted in other fields. The school describes the research as using control groups and variables to determine what instructional methods work better than others (SMU, 2005). Education Secretary Margaret Spellings, an architect of No Child Left Behind, and an advisor to President Bush since he was governor of Texas states, “SMU is approaching this in a formalized overt way (NY Times 2005, p.19).” Her appearance at the commemorative opening of the school marked the support for the program by the Bush administration. Spellings further stated, “SMU’s program will replace the good intentions, abstract philosophies, untested theories, and sometimes harmful fads” (p.19). Deborah Stipek, current dean of the school of education at Stanford is a vocal critic (NY Times, 2005, p.19). Stipek agrees with those who view the dependence on testing as too narrow and rigid. She states, “It is very dangerous and wrong headed, issues like social development are really important as well” (p.19). Other critics remain skeptical and defend existing schools of education that use science based research and blend it with other approaches to give their
graduates a menu of tools to use in the classroom. Kenneth M. Zeichner (2005), associate dean of the school of education at University of Wisconsin and co-author of a report for American Educational Research Association, concludes that all sorts of research is necessary to determine what works best in the classroom. Zeichner described teaching to No Child Left Behind as a “focus on low level learners devised only to improve test scores,” adding: “not that it’s not important. But, all kids need high quality education” (NY Times 2005, p.19). Margaret Spellings disagrees with the critics and emphasizes “major overhauls of teacher preparation and accountability as the only way to improve learning” (NY Times 2005, p.19). A seemingly straightforward analysis may not really indicate student progress or provide educators with enough information to justify the high stakes ranking of schools and teachers (Alleman & Brophy, 2002). The complexity of using high stakes scores and exacting meaning from them cannot be overstated (Folmer, 2001). And, at some point, educators must engage in a critique of the goals of high stakes testing and question adherence to that mission (Kohn, 2000a & 2000b). Educators cannot allow their students to become collateral damage in the debate.

Summary

The literature related to high stakes testing is fraught with concerns, controversies, and inconsistencies. Aspects of the literature that relate directly to social studies high stakes testing are even more obscure and lacking in direction due to few years of testing experience, lack of emphasis on this subject area, and relatively little performance data available for use in analysis, prediction, assessment of programs, and overall guidance. With Texas being one of the few states using high stakes tests for social studies, this study was intended to be a beginning of analysis of the impact of strategies, program and curricular change, and student performance in the area of social studies high stakes testing to fill the social studies testing knowledge gap.
CHAPTER 3
RESEARCH DESIGN

Overview, Statement of the Problem, Design

The class of 2005 in Texas high schools was the first to take the new TAKS test as a requirement for high school graduation. The test results for that group will serve as a prototype for future analysis of student achievement and test modifications. One thing that will not change is the addition of social studies as a core area of assessment. Until the 2005 test, the four core areas of the TAKS have not shared the same history in terms of test preparation and analysis, nor have they been viewed as parallel in importance by TEA. For that reason, the area of social studies instruction and testing begs for refinement and study (Alleman & Brophy, 2000).

In this chapter, an overview and purpose of the study is discussed. The problem is stated, as is the rationale for the quasi-experimental design. The sampling methodology, description of the groups, and subpopulations of interest are discussed. The chapter will cover the assignment to program or comparison group, the description of the data sources, the data collection procedures, the identification of variables for analysis, and data analysis. I have given some attention to the unique concerns surrounding the analysis as a social problem for which social science research is relevant and real world (Judd, Smith, & Kidder, 2001). The methods of analysis will be detailed in the chapter. The following Research and Study Timeline (Figure 1) outlines the testing timeline, availability of data, and other critical elements of the research study process.
Figure 1. Research and study timeline.

1999  Texas Education Code, Senate Bill 103 Mandates TAKS.

2000  Social Studies Center established through TEA funding.

2001  TEKS/Tools Grants available to qualifying districts through SSC.

Decatur ISD applied for and received grant; I received district approval for
proposal for study and implementation of grant requirements; planning for staff
training and curriculum cohesion began.

2002  I used a locally-made pretest for placement and other stipulations of grant (April
2002); training and curriculum adjustments ongoing; I served on state test
development committee.

Late 2002 to spring 2003, implementation of grant requirements/treatment began for
Decatur and other schools in the CCG program schools.

2003  February first state testing occurred; 10\textsuperscript{th} grade social studies TAKS scores;
application of training and curriculum adjustments intensify; particular attention
to at-risk subgroups; grant follow-up is ongoing for all participating schools.

2004  April-11\textsuperscript{th} grade TAKS testing of prototype class of 2005, which were the first
Exit social studies tests in Texas.

2004  Summer- availability of scaled scores for analysis.

2004  Fall- proposal for research using scaled scores of program group and comparison
group for data analysis. Decatur sample school scores from three years available
for analysis.

2005  First Texas TAKS diploma-dependent students graduated in May-June.
Statement of the Problem

Schools receiving the TEKS/Tools Grant Program assumed their changes to the structure of the curriculum and instructional delivery would result in some improvement in student scores on the social studies TAKS test (Social Studies Center, 2001). If that assumption was valid, then there would be an improvement in the social studies TAKS test scores of schools participating in the TEKS/Tools Grant Program as compared to the social studies TAKS test scores of the non-participating schools. There would be an improvement in the Hispanic and low SES subpopulations social studies TAKS test scores in the schools participating in the TEKS/Tools Grant Program as compared to the Hispanic and low SES subpopulations social studies TAKS test scores of the non-participating schools. The program target school (Decatur) students would show improvement from grade 9 to 10 to 11 on the social studies TAKS test scores.

This study used a quasi-experimental design to investigate the first TAKS social studies scores available for analysis in Texas. This study used a nonrandomized control group design. Decatur, the target sample school, was part of a larger group of schools with similar demographic and funding characteristics. During the period of data collection and analysis, every school in Texas had a unique comparable group of schools available for purposes of study and analysis (See below for an explanation of campus comparison group (CCGs)). Because all schools in Texas had the opportunity to be part of the TEKS/Tools Grant program (by virtue of application and compliance with grant requirements in 2001-2002) member schools of this unique comparison group were considered for participation in this study ($n = 41$) (Abelson, 1995). After eliminating some schools based on inability to meet study criteria (See explanation below), the remaining CCG schools ($n = 37$) were assigned to the experimental group ($n = 8$) if it was determined they had participated fully in the treatment which was the TEKS/Tools Grant
Program training and follow-up (Isaac, 1981) or they were assigned to the control group \( n = 29 \). For purposes of this study, which uses a quasi-experimental design without random assignment, the group participating is referenced as the program group and the non-participants in the TEKS/Tools Program are referenced as the comparison group. The use of this terminology is to acknowledge the quasi-experimental nature of the design and avoid any claim of this study as a true experiment. The use of intact school groups was critical to the study.

Availability of a comparison group provided a group of subjects who did not receive the treatment but in all other respects of the TAKS administration process were well matched to the program group. This design allowed examination of the effects of the treatment (adherence to the TEKS/Tools Grant Program training and curriculum emphasis) itself. The CCG consisted of Decatur High School, as the target school, and schools identified by the formula from Texas Education Agency for CCGs. In general, the CCG is a group of schools considered well matched under multiple criteria established by TEA (i.e., ethnicity, achievement, funding, the number of economically disadvantaged students reported to the state based on free and reduced lunch figures, the percentage of students identified with limited English proficiency, etc., as detailed in Appendix C). Even with the closeness of the match across schools, individual differences and limitations necessitated adjustments and certain schools were not included in this study. For example, this study did not include member schools with glaring demographic or other inconsistencies when juxtaposed with the demographics of the Decatur target school. For purposes of analysis, a homogenous match of schools-subjects resulted (Oakes, M. 1986). (See Appendix C for a full explanation of state CCG formation.)

Various statistical analysis techniques were used to study the data generated. In approaching the decision as to technique and analysis, I attempted to follow the four guidelines
recommended by D. R. Krathwohl (1998) for those engaging in educational and social science research: 1) to consider this study as a social process through which findings become accepted as knowledge; 2) to comprehend how the criteria by which this research is judged flow from the knowledge development process; 3) to be able to read, analyze, understand and evaluate this research using a variety of approaches; and 4) to understand the importance of handling this research responsibly and ethically and value the choices involved therein. This approach seems to encourage a multi-faceted look at the emerging data and encourage interpretations that are fluid in their practical meaning. The ethical considerations and implications alone are substantial when we consider the population we are studying. With those goals in mind, I used gain scores, ANCOVA, and repeated measures analysis to examine the overall score gains of the program-comparison groups, to examine the scores of program-comparison groups for significant improvement and various effects, and to identify interactions, trends or emerging questions for study. The inclusion of analysis that focused on the subpopulations of concern as variables seemed critical after reviewing the literature and the general concern of educators regarding at risk groups (Langfield, Thurlow, & Scott, 1997; Ohanian, 1999; Serow, 1984).

The methodology included two separate analyses. One set of analyses included all members of the program-comparison groups. The other set of analyses included only the data from Decatur High School, the target school within the CCG, which will be referred to as the Decatur sample school in that analysis. The Decatur target school is the nucleus school from which the CCG was developed and is also the school where the curriculum development originated. This second analysis differed because additional data were available for this school only. Specifically, while the program-comparison groups’ schools had mean test scores available for grades 10 and 11, the Decatur sample school also had data from grade 9. While all other data
consist of school mean scores only, The Decatur sample school offered much richer and more specific information in the form of 122 individual student scores for three test administrations (grades 9, 10, and 11). Note: All schools (subjects) have scaled mean scores for gains score analysis, ANCOVA, and RM using two measures (10 and 11); this includes the subject school, which is the Decatur target school. The Decatur sample is also used for a separate trend repeated measures analysis as it had the advantage of 122 individual scores for three measures (9, 10, 11).

Participants: The Decatur Sample School

The Decatur sample group consisted of the class of 2005 at Decatur High School. The original freshmen class numbered approximately 200. With attrition and absence, however, the final sample for this study contained 122 students \((n = 122)\), all of whom were present for the three days of testing. The class demographic profile included 75% Anglo, 22% Hispanic, with African American and “other” comprising less than 3% (using the stated TEA guidelines, TEA, 2004). Approximately 23% of the students were identified as low socioeconomic status based on their qualification for reduced or free lunch. (That number may be underreported slightly due to failure of parents and staff to encourage application and paperwork for qualification.)

The subpopulations of interest, Hispanic and low socioeconomic, are the subpopulations that are growing in Texas and in Decatur based on demographic reports from the State of Texas Comptroller’s Office (Sharp, 2002 update). This growth trend is of concern for school districts in Texas as the AEIS reports have confirmed the consistently lower TAAS and end-of-course scores from those populations since reporting procedures began to track the groups’ performance data (TEA reports 1990 to 2003). This demographic trend prediction is paralleled across the United States (Pallas, Natriello, & McDill, 1989; US Census Bureau 2004-2005).
Description of the Program and Comparison Groups

The program (participants in TEKS/Tools Program) and comparison (non-participants) groups are described together for the sake of ease. The program group (which includes the Decatur sample group) and the comparison group were derived from the schools that made up the Decatur CCG. Texas Education Agency explains the CCG in this way:

Each campus has a unique comparison group of 40 other campuses in the state that closely match the target school on a number of characteristics. Comparison groups are recreated each year to account for changes in demographics that may occur. They are used for all group statistics reported on campus AEIS reports and the School Report Cards. (http://www.tea.state.tx.us/perfreport/account/2002/manual/sec05.html)

The target or sample school in this study was Decatur High School. The CCG used for purposes of the study was established in 2002-2003, as that was the first year the social studies TAKS test was given for accountability purposes. For this study, only self-contained 9-12 schools were included. Interestingly, the CCG composition did not change throughout the course of the study (see Figure 2). The use of the CCG allowed me to employ a form of matched-subjects design in which the subjects are matched on one or more relevant characteristics, reducing between groups variability. Comparing schools that are highly dissimilar reduces the practicality of generalizing programs’ effects.

Forty-one schools comprised the CCG; of those, eight were determined to have participated fully in the TEKS/Tools Grant Program in terms of curriculum design, teacher training and implementation. Those became the treatment group, known as the program group in the quasi-experimental design. Four schools were deleted from the study due to errors in score reporting and problems with score tracking from 10th to 11th grade. The remaining 29 schools were assigned to the control group, known as the comparison group in the study design.
Figure 2. Decatur campus comparison group 2002-03 demographics.

<table>
<thead>
<tr>
<th>CAMPUS</th>
<th>DISTRICT</th>
<th>WHITE</th>
<th>HISP</th>
<th>ECON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smithville HS</td>
<td>Smithville ISD</td>
<td>72.2</td>
<td>15.4</td>
<td>29.6</td>
</tr>
<tr>
<td>Belton HS</td>
<td>Belton ISD</td>
<td>73.3</td>
<td>20.7</td>
<td>26.4</td>
</tr>
<tr>
<td>A &amp; M Cons HS</td>
<td>College Station ISD</td>
<td>73.4</td>
<td>08.9</td>
<td>16.7</td>
</tr>
<tr>
<td>Marble Falls HS</td>
<td>Marble Falls ISD</td>
<td>73.8</td>
<td>23.3</td>
<td>29.7</td>
</tr>
<tr>
<td>Farmersville HS</td>
<td>Farmersville ISD</td>
<td>75.6</td>
<td>19.0</td>
<td>27.2</td>
</tr>
<tr>
<td>Vines HS</td>
<td>Plano ISD</td>
<td>75.6</td>
<td>10.3</td>
<td>11.1</td>
</tr>
<tr>
<td>Shepton HS</td>
<td>Plano ISD</td>
<td>72.6</td>
<td>05.6</td>
<td>06.4</td>
</tr>
<tr>
<td>Columbus HS</td>
<td>Columbus ISD</td>
<td>75.5</td>
<td>14.7</td>
<td>27.4</td>
</tr>
<tr>
<td>Pearce HS</td>
<td>Richardson ISD</td>
<td>72.4</td>
<td>16.4</td>
<td>11.6</td>
</tr>
<tr>
<td>Palmer HS</td>
<td>Palmer ISD</td>
<td>74.2</td>
<td>24.1</td>
<td>22.3</td>
</tr>
<tr>
<td>Red Oak HS</td>
<td>Red Oak ISD</td>
<td>77.9</td>
<td>14.1</td>
<td>10.5</td>
</tr>
<tr>
<td>Maypearl HS</td>
<td>Maypearl ISD</td>
<td>77.5</td>
<td>18.6</td>
<td>22.9</td>
</tr>
<tr>
<td>La Grange HS</td>
<td>La Grange ISD</td>
<td>76.0</td>
<td>14.9</td>
<td>23.8</td>
</tr>
<tr>
<td>Clear Creek HS</td>
<td>Clear Creek ISD</td>
<td>1.6</td>
<td>15.5</td>
<td>08.4</td>
</tr>
<tr>
<td>Pampa HS</td>
<td>Pampa ISD</td>
<td>75.7</td>
<td>19.5</td>
<td>25.8</td>
</tr>
<tr>
<td>Pine Tree HS</td>
<td>Pine Tree ISD</td>
<td>75.1</td>
<td>12.1</td>
<td>18.1</td>
</tr>
<tr>
<td>Marion HS</td>
<td>Marion ISD</td>
<td>72.0</td>
<td>25.3</td>
<td>22.9</td>
</tr>
<tr>
<td>Katy HS</td>
<td>Katy ISD</td>
<td>75.0</td>
<td>17.5</td>
<td>11.6</td>
</tr>
<tr>
<td>Klein Oak HS</td>
<td>Klein ISD</td>
<td>77.4</td>
<td>13.8</td>
<td>10.5</td>
</tr>
<tr>
<td>Klein Collins HS</td>
<td>Klein ISD</td>
<td>74.9</td>
<td>13.0</td>
<td>07.2</td>
</tr>
<tr>
<td>Sulphur Springs HS</td>
<td>Sulphur Springs ISD</td>
<td>75.9</td>
<td>12.1</td>
<td>22.7</td>
</tr>
<tr>
<td>Alvarado HS</td>
<td>Alvarado ISD</td>
<td>78.2</td>
<td>16.7</td>
<td>35.4</td>
</tr>
<tr>
<td>Lampasas HS</td>
<td>Lampasas ISD</td>
<td>78.1</td>
<td>16.6</td>
<td>27.0</td>
</tr>
<tr>
<td>Dayton HS</td>
<td>Dayton ISD</td>
<td>77.6</td>
<td>10.8</td>
<td>27.8</td>
</tr>
<tr>
<td>Groesbeck HS</td>
<td>Groesbeck ISD</td>
<td>75.1</td>
<td>12.1</td>
<td>33.8</td>
</tr>
<tr>
<td>Lubbock Cooper HS</td>
<td>Lubbock Cooper ISD</td>
<td>71.9</td>
<td>26.8</td>
<td>29.7</td>
</tr>
<tr>
<td>Shallowater HS</td>
<td>Shallowater ISD</td>
<td>72.2</td>
<td>26.9</td>
<td>24.2</td>
</tr>
<tr>
<td>Greenwood HS</td>
<td>Greenwood ISD</td>
<td>77.6</td>
<td>20.5</td>
<td>19.7</td>
</tr>
<tr>
<td>Willis HS</td>
<td>Willis ISD</td>
<td>75.3</td>
<td>16.2</td>
<td>30.5</td>
</tr>
<tr>
<td>Breckenridge HS</td>
<td>Breckenridge ISD</td>
<td>74.9</td>
<td>22.2</td>
<td>22.6</td>
</tr>
<tr>
<td>Martin HS</td>
<td>Arlington ISD</td>
<td>78.0</td>
<td>06.7</td>
<td>05.9</td>
</tr>
<tr>
<td>Bell HS</td>
<td>HEB ISD</td>
<td>74.7</td>
<td>11.0</td>
<td>20.3</td>
</tr>
<tr>
<td>Boswell HS</td>
<td>Eagle Mt-Saginaw</td>
<td>73.9</td>
<td>16.8</td>
<td>12.2</td>
</tr>
<tr>
<td>Grape Creek HS</td>
<td>Grape Creek ISD</td>
<td>74.1</td>
<td>25.3</td>
<td>34.5</td>
</tr>
<tr>
<td>Georgetown HS</td>
<td>Georgetown ISD</td>
<td>75.2</td>
<td>20.7</td>
<td>21.7</td>
</tr>
<tr>
<td>Georgetown 9th HS</td>
<td>Georgetown ISD</td>
<td>73.8</td>
<td>23.1</td>
<td>26.5</td>
</tr>
<tr>
<td>Hutto HS</td>
<td>Hutto ISD</td>
<td>74.0</td>
<td>17.5</td>
<td>16.1</td>
</tr>
<tr>
<td>Westwood HS</td>
<td>Round Rock ISD</td>
<td>75.1</td>
<td>09.6</td>
<td>05.1</td>
</tr>
<tr>
<td>Leander HS</td>
<td>Leander ISD</td>
<td>75.0</td>
<td>17.0</td>
<td>19.1</td>
</tr>
<tr>
<td>Wink HS</td>
<td>Wink-Loving ISD</td>
<td>72.3</td>
<td>25.8</td>
<td>27.7</td>
</tr>
<tr>
<td>Decatur HS</td>
<td>Decatur ISD</td>
<td>74.9</td>
<td>21.9</td>
<td>20.5</td>
</tr>
<tr>
<td>Group Average</td>
<td></td>
<td>75.0</td>
<td>14.9</td>
<td>15.7</td>
</tr>
</tbody>
</table>

Note: T = Treatment Groups; D = Deleted (List is in order of formula match to target school.)
Three sources of verification were used in the process of determining the final group of schools to be used in the study. Karen Wiggins, Director of the TEA Social Studies Center, was responsible for the approval of grant applications, verification of compliance with the replication proposal conditions, and the funding to districts of grant monies. Bill Shuttlesworth, Social Studies Specialist Region VI Education Service Center, made the second verification. By TEA definition, the education service centers (ESCs) provide state leadership for special education related decentralized functions. Region VI Education Service Center was the assigned field structure for professional development and other duties associated with the TEKS/Tools Grant Program. I made the third verification of TEKS/Tools Grant Program participation and compliance by contacting each school in the CCG for the target school, Decatur High School. This verification process was documented through several interview formats with administrators, teachers, and appropriate coordinators for the districts. I asked the following questions and compiled narrative regarding the school responses. The questions included:

1. Describe the school’s reasons for participating in the program or not.

2. Describe the school’s staffing prior to, during, and immediately after the time frame in question (turnover, attrition, changes in placement for teachers). Describe any additional training for staff.

3. Describe the emphasis and support by administrative teams.

4. Detail any other factors that you (the contact coordinator) think affected the students in their testing environment.

The following schools were deleted from the study: (the primary reason is noted for each.) Wink High School in the Wink Independent School District was deleted due to incomplete record keeping. Vines High School and Shepton High School, both in the Plano Independent School District, were deleted because the students assigned to those schools were monitored for only one year. Georgetown 9th Grade School was deleted because that school had only 9th grade
students with no follow-up plan. The original number of 41 schools was reduced to 37 with 8 schools (the treatment group) participating in the TEKS/Tools Grant Program, the intervention in question, and 29 schools being non-participants. By virtue of that non-participation, those 29 schools were assigned to the control group for the quasi-experimental study.

Procedures

Decatur Sample Methodology

The Decatur sample group consisted originally of 200 students from the class of 2005 at Decatur High School. Decatur was an availability sample that provided a window to three years of access and data unavailable in the comparison schools, whether program participants or not. The use of this sample for measuring three years of score improvement provided an examination of more information within a larger treatment sample. Human subject forms and a general proposal for the study were completed and approved by the district superintendent and school board. Due to attrition for various reasons, the actual Decatur sample group was reduced to fewer than 150 students. The number of students from the Decatur sample group who tested in 9th, 10th, and 11th grade was 122 students ($n = 122$).

Curriculum and Treatment Design

The treatment used in this study resulted from a two-year grant obtained from the Texas Education Agency, the TEKS/Tools Grant Program. The TEKS/Tools Grant Program (Appendix B) was established after the Legislature passed Senate Bill 103 to assist in the transition to the new testing and accountability system with social studies as a tested area. The
training and implementation focused on professional development that emphasized established best practices with particular attention to several areas (TEA, 2001):

- Data driven decision-making
- Elementary social studies: building a foundation
- United States history: developing a sense of time and place
- World history and world geography: connecting concepts
- Economics, government, and culture: 21st century citizenship
- Integrating content and skills instruction (see Appendix B for a thorough explanation of these six areas)

Those conceptual frames were stressed in partnership with both horizontal and vertical alignment of the curriculum, and a delivery system that emphasized teaching to enhance critical thinking and inquiry approaches. The tight alignment of the new TAKS test to the TEKS cannot be overstated and the use of the TEKS to build classroom curriculum was a primary foundation for the training and the follow-up program. The close alignment of the Texas curriculum, the TEKS, and the test, the TAKS, was the primary focus of all professional development and preparation. The Decatur Independent School District social studies teachers participated in curriculum development and implementation, as did instructors in other districts who were awarded a TEKS/Tools Grant (the program group receiving the treatment). Curriculum design was ongoing; however, instructors began implementing treatment-based teaching late in the 2002-2003 school year. In 2003, the grant was extended to bridge the year 2003-2004. (Only schools receiving this extension were considered for the comparison group.) The grant ended in February 2004. Texas Education Agency states: “The TEKS are the framework for TAKS assessment objectives. A thorough understanding of the TEKS is essential for student success in
the classroom, on state assessment, and as reflected in the state accountability system (TEA 2005).” The TEKS became effective in all content areas on September 1, 1998. However, it was apparent that many districts had not had recent curriculum alignments or audits to reflect the emphasis on the state curriculum guidelines, known as the TEKS. This specific focus became the intent of the TEKS/Tools Training Program with the grant monies that were attached to the training. The necessity of schools to align and teach the TEKS was a stated goal of the overall state program and a specific goal of the Decatur target school. Curriculum planning was a major part of the TEKS/Tools Grant Program. All participant schools in the CCG were supposed to apply the planning, training, guidelines, and instructional strategies to their curriculum and instructional social studies design.

Decatur approached the change as they were instructed by TEA through the ESC representatives. First, the district audited the existing social studies curriculum as to course offerings K through 12, content in those offerings (using the TEKS guidelines throughout), and adjustments were made. Every social studies teacher’s lesson plans and the district curriculum were examined for detail. Some plans were in alignment; many were not. Adjustments were made and the TEKS were used to formulate new curriculum as necessary. Staff development was an integral part of the program. All teachers who trained initially with the TEA and ESC representatives were obligated under the terms of the grant to replicate the training for any coworkers who dealt in any way with related subject areas and any district staff member who may not have been part of the initial training teams. Funding was dependent on the adherence to the conditions and terms of the grant, which specifically demanded curriculum alignment and subsequent training. The specific areas covered by the TEKS/Tools Grant Training Program and the target school manner for adherence follows (as cited in assumptions, the target school
proceeded in the manner that all compliant schools were assumed to have proceeded to address their grant requirements as a stipulation for funding).

“Data driven decision-making” was one of the six main components. Decatur began to access any social studies test data on students. This assessment included the previous TAAS tests scores and the end-of-course tests scores. It became apparent that neither one of the tests was parallel to the projected new design of the TAKS tests in rigor or format. Decatur decided to administer a local pretest that I had designed. This test approximated the TAKS format and served as a pilot study of the student behaviors and abilities for the study. A description of the Decatur baseline test development and administration follows the discussion of the six areas of program emphasis. Data collection, disaggregation, and analysis of student progress became an ongoing procedure with a systematic approach that all teachers were required to follow (NCREL 2005).

The second component of the program was “elementary social studies: building a foundation.” This concept required all teachers, including those at elementary levels, to investigate best practices for the subject area as well as grade level. The teachers had to agree on a consistent use of terminology so that concepts built from year to year. (Example: if the TEKS used the term *urbanization* from K-12, then all teachers were expected to use that specific term, not one of their personal preference, even if it had the same meaning.) Secondary teachers worked with elementary teachers to refine this process and other areas that needed alignment. (Example: if the TEKS and TAKS were going to cover Benjamin Franklin, then the elementary teacher could use that historic figure for a unit of study rather than Johnny Appleseed, a non TEKS-non TAKS mythical figure). Using Piaget’s theories about child development and readiness for concepts, the teachers approached the concepts and themes in a cautiously
constructed scaffold where ideas that are age appropriate are introduced and reinforced (Grant 2001). Then the elaboration and extension of those concepts are built upon the foundation (Piaget 1983).

The next three components are listed together as they are content driven. They are: “United States history: developing a sense of time and place”; “world history and world geography: connecting concepts”; and “economics, government, and culture: 21st century citizenship.” These content strands compose the nucleus of the content for every level K-12 in the TEKS. The TAKS test has questions from each of the three large umbrella areas. The number of questions from each varies slightly but is linked by connecting concepts. The cross content approach is one that is an established method of teaching history and is encouraged by the National Council for Social Studies (NCSS 2005) and encouraged by current social studies researchers (Pahl 2003, Crocco 2003, Libresco 2004).

“Integrating content and skills instruction” was the sixth area of program emphasis. There are many reasons for implementing integrated units, for example, aiding students in making interconnections among subject areas, learning generic skills (e.g., problem solving, problem posing, integrative skills, collaborative skills, interpersonal skills), or finding personally meaningful entrance points to the curriculum (Barab 1998). Using Bloom’s taxonomy to construct and refine questioning strategies (Anderson & Krathwohl 2000); Vygotsky’s zone of proximal development and socio-cultural origins of mental functioning to connect content to context (Cole 2003); and Bruner’s theories on transfer of learning to reinforce concepts, scaffold and connect learning goals (Bruner 1977), the teachers and program adherents refined the methods of consistent instruction. This focus on the “construction” of a rich curriculum and its delivery through an instruction process that was aware of the learner’s readiness and needs and
used a shared vocabulary and content approach became the core of the application of the TEKS/Tools Grant Program within the target district. All participating schools had instruction and expectations from TEA regarding this approach. The replication of this process was a criterion for funding in the program schools.

Decatur Baseline Test Development

In the interest of determining student improvement, the Decatur sample group was administered a pretest in April 2002. While no complete test was available from TEA that could be used to collect baseline data, 13 sample social studies TAKS questions were. I had served on the TEA teacher committee (2001-2002) to review the bank of possible questions that were projected for the 10th grade test and the 11th grade exit social studies TAKS test (see Appendix D for test development process). Using the levels of difficulty, the domains and criteria, and the overall format for the actual tests, a baseline test was formulated. I took the 13 samples and embedded them in the test that was constructed using the stated criteria and domains for testing as outlined by Texas Education Agency (Social Studies Center, 2002). The baseline test was administered to the Decatur students in April 2002. The baseline test scores were used by Decatur teachers in the TEKS/Tools Grant to inform their professional development and curriculum planning. These scores were also used for Decatur’s profile in the repeated measures analysis described later. A description of the reliability and validity of the baseline test is discussed in chapters 4 and 5.

The instructors regarded the period of time between the baseline and the first official administration of the 10th grade social studies TAKS in February 2003 (two months earlier than
the traditional April state test dates). Instructors implemented the treatment during the 2003-2004 school year. The class of 2005 was tested on the 11th grade TAKS exit test in April 2004.

Description of Data Sources

The other data sources in addition to the baseline test described above were the TAKS social studies tests administered in February 2003 and April 2004. Samples of those tests may be found online as released tests by TEA (TEA, 2005). A thorough description of the test construction criteria and processes are included in chapter 2 and Appendix D. The Texas Education Agency analysis of the tests indicated a high degree of internal validity (Texas Education Agency, 2003-2004). Using a public information request, the mean scores for each school subpopulation in the CCG from TEA were obtained. The Decatur sample scores were obtained through the following testing procedures.

Testing Procedures

The baseline local test was administered to the Decatur students, the class of 2005 at Decatur High School in April 2002. The test conditions purposely simulated actual TAKS testing conditions by administering the test during the formal TAAS testing of the class of 2004. To further convey the importance of the test to the students, the tests were counted as an actual grade within their classes. Conditions for testing were carefully controlled to simulate the actual TAAS/TAKS environment. Such things as actual rooms, lighting, time framework, and even the same juice-muffin snack at 8:00 am were included. The entire high school staff participated, some administering the TAAS and some administering the sample TAKS. All staff was trained by the counselors with the same training given for TAAS and sample TAKS. All staff
completed the oath for confidentiality and security issues. Students were unaware that the local test was not part of TAAS.

The local tests scores were later used to measure any achievement when compared to the actual 2003 10th grade TAKS social studies test scores. The 10th grade TAKS social studies test was given in February 2003. After the 2003 scores were retrieved, the local test was checked for validity and reliability accuracy. Of particular interest was the predictive validity of the local test. In analyzing the scores and any trends, the Decatur subpopulations of interest were also monitored and were included in the preliminary analysis.

To complete the full study, the scores on the 2004 test, which was administered in April 2004, were used further to measure student achievement. The 2004 scores determined whether a student passed the social studies TAKS for the purpose of receiving a diploma. The CCG program (experimental) group schools and the CCG comparison (control) group schools were tested using the same Texas Education Agency generated instructions, security measures and procedures, time framework, and staff test administration instruction and monitor safeguards. Any school with questionable methods for testing or reporting was deleted from the study. The testing procedures and the test instruments determined by Texas Education Agency were parallel for both the treatment group schools and the control group schools. Therefore, I am confident that the procedures and methods for both groups were as precise as a school setting would allow.

Collection of Data

The Decatur sample data were collected for the available three years with the cooperation of the Decatur Independent School District. I was one of the Decatur team of monitors for the state tests. As data became available, I gathered the scores for analysis. Assisted by the social
studies department at Decatur High School, I scored the 9th grade local test. All subjects were anonymous for reporting purposes.

The collection of the program (experiment) group schools’ data and the comparison (control) group schools’ data was accomplished with the assistance of the Public Information Act of 1999, the cooperation of the Texas Education Agency, and the cooperation of the cohort schools. Test scores were collected for all subjects and for defined subpopulations. The defined subpopulations of interest included the Hispanic subpopulation and the low socioeconomic subpopulation. Texas Education Agency representatives did the scoring for all CCG schools’ 10th and 11th grade scores, both program group and comparison. I collected all available scores related to the TAKS tests for the years 2003 and 2004. The scores are represented by a scaled mean of the tested subjects for that school. The scaled mean score is the best available score for use in comparing school scores. They are consistent from school to school. They offer a score that is suited to gains score analysis and ANCOVA analysis. The social studies scaled score range was from 1406 to 2759. The passing score was set by TEA at 2033 and a score of 2400 or better was considered commended. The standards for passing and commended scores are determined by TEA after the tests are administered and scored. Other demographic differentiation by scaled mean score was available for study and was collected for preliminary analysis and understanding. Data on other subpopulations were gathered for future study. I collected, recorded, coded, and formatted all data for use with Statistical Package for Social Sciences Program, version 13.
Statistical Analysis

In addition to the social studies scores, I gathered the other core tested area scores: math, language arts, and science. The inclusion of other core area scores allowed the exploration of correlations between areas of testing, again, for future study. For initial data analysis, I used an exploratory approach (Borg, Gall, & Gall. 1998 & 2002), which has been very much emphasized by recent innovators in the field of social science research as a way of discerning important relationship across content areas as they relate to the social sciences. (Cohen, 1990 & 1994). Such an emphasis is particularly important with data that have never been accessed prior to the first generation of tests (Kennedy, 2003). Peter Kennedy’s suggestions for initial data analysis (IDA) stresses casting a wide net for data analysis and emphasizing the initial phase. The process includes an inspection of the behavior of as many variables as possible, outlier analysis and awareness of irregularities, descriptive and graphical output (getting a feel for the structure or patterns of the data), and possible model reselection or modification. Although criticized by some as too investigative, defenders view the approach as having a common sense perspective and supportive of the notion that an approximate answer to the right question is worth more than a precise answer to the wrong one (Kennedy, 2005; p. 98). An obvious lack of control regarding the intervention strategies and when and with what force they were applied became a driving factor in the analysis and the rationale for the use of several methods of analysis to remedy the intervention’s unevenness. In that regard, I gathered, formatted and analyzed much more data than this particular study will reveal in its findings and evaluation. I also expanded the repeated measures mixed design model, because certain interactions would have been unknown, overlooked and unreported. For purposes of the quasi-experimental analysis (which included gains score analysis, ANCOVA, and repeated measures using both within subjects and between
subjects analysis) only the social studies scores were used. For purposes of the Decatur sample repeated measures analysis, only the social studies scores were used.

**Decatur Sample Group Analyses**

For the analysis of the data from the Decatur sample group only, individual TAKS social studies scores were analyzed according to the procedures for a repeated measures design using 9$^{\text{th}}$, 10$^{\text{th}}$, and 11$^{\text{th}}$ grade scores. For comparisons between the program and comparison groups, the Decatur mean score was used, as only mean scores were available for the CCG schools. For the between group comparisons, 10$^{\text{th}}$ and 11$^{\text{th}}$ grade TAKS scores were used. The analysis of the Decatur sample group used in this study was chosen to determine changes in student performance over the three-year period for which data were available. The analysis of the Decatur sample, a subset of the program group, used repeated measures, a $t$ test with three repeated measures (the three years of local test scores), analysis of variance or ANOVA. The use of the Decatur trend analysis is exploratory as it was used to investigate an area about which little information exists. The profile provided three years of data and included the use of a local pretest as a form of pilot study serving as a trial run of the experiment and as a vehicle for gaining more information and formative program evaluation (Stufflebeam, 1971).

**Campus Comparison Group Analyses**

For purposes of analysis with the CCGs, the 10$^{\text{th}}$ grade TAKS social studies test served as the pretest (independent variable) and the 11$^{\text{th}}$ TAKS social studies test served as the posttest (dependent variable) and they were used as measure one and measure two of the within subjects measures. The February 2003 and 11$^{\text{th}}$ grade exit test administered in April 2004 were used as
the pre and post test scores for the ANCOVA analysis of the treatment schools and the control schools.

The analysis of the CCG in this study is descriptive, parametric, and inferential. I used several forms of analysis (gains score analysis, ANCOVA) as corroborative evidence for the validity of the research findings (Leedy, 2001). The procedures and measures were used to make inferences about the population characteristics (the high school students of Texas who will take an exit test in social studies as a graduation requirement) from the program and comparison groups identified and drawn from that population, the CCG. The statistical analysis of the CCG groups, both program and comparison groups, used a mixed factorial design. The factorial design includes both between and within subjects variables. This type of mixed design, which is particularly powerful, is a pre-post-control design in which all subjects (all CCG schools) were given a pre-test (10th grade TAKS) and a post-test (11th grade TAKS), that together serve as a within-subjects factor (test). Those scores were also used as two measures in the repeated measures analysis of the CCG schools. A repeated measures analysis of variance was used on the CCG schools both program and comparison groups with the 10th and 11th grade scores as a 2 x 2 ANOVA with time (pretest vs. posttest) as a within-subjects factor and treatment (treatment vs. control) as a between subjects factor (Borg, Gall & Gall, 2002).

I initially used analysis of gain scores to examine the data. Gain scores were used to obtain a preliminary understanding of the CCG scores. That procedure was inconclusive and had inherent hazards including ceiling effects, regression toward the mean, assumption of equal intervals, and a possible low reliability (Borg, Gall, & Gall, 2002). To attempt to address those hazards, I used ANCOVA for the analysis of data. Analysis of covariance allowed examination of the differences among means while controlling for the effects of the 10th grade scores
correlated with the dependent variable (the 11th grade or exit scores). An acknowledged problem is the lack of random assignment for the ANCOVA analysis. Implications of a Type II error will be discussed in chapter four. For ANCOVA purposes, the testing of the null hypothesis was done at the .05 significance level.

The subpopulations of interest were analyzed as necessary and with parallel analyses or as variables as appropriate. I included the Decatur sample because that school was the only one of the CCG that had three years of scores for study. Established researchers in the field of social sciences have consistently encouraged the inclusion of corroborative data and analysis whenever possible, especially when the area being studied is new to the research arena (Kennedy, 2003, also, Borg, Gall, & Gall 2002). The following discussion outlines the rationale and considerations for using a varied approach for analysis.

Repeated Measures Design

I used a repeated measures design to analyze the Decatur schools’ scores. Gain scores and ANCOVA are useful in confirming (or not) the null hypothesis; however, realizing that there is meaning beyond statistical significance, the repeated measures design offered some distinct advantages including more information about the effect of the treatment. Christopher Chatfield (1991), in Avoiding Statistical Pitfalls, cautions about the very concerns expressed in this narrative. In treating the scores to a repeated measures ANOVA, the study became a quasi-experiment with multiple scores for each case (or school). The repeated measures or within subjects design reduces the nonsystematic variance (the error that is not under experimental control), removes the variance due to individual differences, has more sensitivity or power (in the case of my limited access to the scores of every student rather than only to the mean score of
the school), and there is a certain efficiency in the analysis operation (p. 240-252). The repeated measures design is useful for measuring performance on the same variable over time, which fits the need to look at changes in student performance on the TAKS tests before and after a specific treatment, the effects of the TEKS/Tools Grant Training Program. The same subject is measured multiple times under different conditions; the same subjects provide measures/ratings on different characteristics (the independent variables of race, low SES, and gender).

**Research Questions**

1. Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the Comparison (control) group that did not participate?

2. Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between low socioeconomic students in the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the low socioeconomic students in the Comparison (control) group that did not participate?

3. Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between Hispanic students in the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the Hispanic students in the Comparison (control) group that did not participate?

4. Is there a statistically significant linear trend in the TAKS social study means across the 9th, 10th, and 11th grades for the Decatur students only?
Hypothesis

The Problems as Null Hypotheses

H01: The program group participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade exit TAKS test scores as compared to the scores of the comparison group that did not participate in the TEKS/Tools Grant Program (control group). (The study compares the mean scores of each school within the treatment and control groups).

H02: The program group low socioeconomic subpopulation participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade exit TAKS test scores as compared to the scores of the comparison group low socioeconomic subpopulation that did not participate in the TEKS/Tools Grant Program (control group).

H03: The program group Hispanic subpopulation participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade exit TAKS test scores as compared to the scores of the comparison group Hispanic subpopulation that did not participate in the TEKS/Tools Grant Program (control group).

H04: There will be no statistically significant linear trend in the TAKS social study test means across the 9th, 10th, and 11th grades for the Decatur student’s scores only.

Summary

I constructed a quasi-experimental study of parallel schools. The schools were matched using an existing Texas Education Agency formula (Appendix C). Those schools participating in the TEKS/Tools Grant program were identified as Program schools. Non-participant schools
constituted the comparison group. To check for significance differences between groups, I used gains scores, ANCOVA, and a mixed design repeated measures analysis with the program and comparison groups. I used repeated measures and mixed design analysis with a sub-sample of the program group (the Decatur sample group) for which three years of data were available for analyzing the significance of possible test score variation over time. Subpopulations of interest (low socioeconomic and Hispanic) were analyzed using ANCOVA and repeated measures methods with either a parallel analysis or the subpopulations as an independent variable as appropriate.
CHAPTER 4
RESULTS

Schools receiving the TEKS/Tools Grant Program assumed their changes to the structure of the curriculum and instructional delivery would result in some improvement in student scores on the social studies TAKS test (Social Studies Center, 2001). If that assumption was valid, then there would be an improvement in the social studies TAKS test scores of schools participating in the TEKS/Tools Grant Program as compared to the social studies TAKS test scores of the non-participating schools. There would be an improvement in the Hispanic and low SES subpopulations social studies TAKS test scores in the schools participating in the TEKS/Tools Grant Program as compared to the Hispanic and low SES subpopulations social studies TAKS test scores of the non-participating schools. The program target school (Decatur sample) students would show improvement from grade 9 to 10 to 11 on the social studies TAKS test scores.

The first research question is: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the comparison (control) group that did not participate? The first research question was answered using TAKS social studies gain scores. The average TAKS social studies gain score difference between the Program group and the comparison group was tested using an independent t-test.

The second research question is: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between low socioeconomic students in the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the low socioeconomic students in the comparison (control) group that did not
participate? The second research question was answered using TAKS adjusted mean social studies scores from low socio-economic groups. The adjusted mean score difference between the program group and the comparison group was tested using analysis of covariance. The pre-test scores were used as a covariate to compute the TAKS adjusted post-test means for the program low SES and comparison low SES groups.

The third research question is: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between Hispanic students in the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the Hispanic students in the comparison (control) group that did not participate? The third research question was answered using TAKS adjusted mean TAKS social studies scores from Hispanic subpopulations in program group versus comparison group. The pre-test scores were used as a covariate to compute the TAKS adjusted post-test means for the Program Hispanic subpopulation and comparison Hispanic subpopulations.

The fourth research question is: Is there a statistically significant linear trend in the TAKS social study means across the 9th, 10th, and 11th grades for the Decatur students only? The fourth research question was answered using TAKS social studies gain scores. The linear trend in TAKS social studies gain scores across 9th, 10th, and 11th grades of the Decatur sample (n = 122) only was tested using repeated measures.

TAKS Social Studies Gain Score Analysis

H01: The program group participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade Exit TAKS test scores as compared to the scores of the comparison group that did not participate in the TEKS/Tools Grant Program (control group). (The study compares the mean scores of each school within the treatment and control groups).
The first research question tested for a statistically significant mean gain score difference between the Program group (treatment) and the Comparison group (control). Table 1 presents the mean TAKS gain scores, standard deviations, and independent t-test results. Gain scores ranged from -17 to +135 (Appendix E). The independent t-test indicated a statistically significant TAKS social studies mean gain score difference. The null hypothesis is therefore rejected.

Table 1

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>8</td>
<td>36.25</td>
<td>26.28</td>
<td>2.01</td>
<td>.05</td>
</tr>
<tr>
<td>Comparison</td>
<td>29</td>
<td>57.30</td>
<td>26.14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Socioeconomic Status

H02: The program group low socioeconomic subpopulation participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade Exit TAKS test scores as compared to the scores of the comparison group low socioeconomic subpopulation that did not participate in the TEKS/Tools Grant Program (control group).

The second research question compared TAKS adjusted mean social studies scores from low socio-economic subpopulations between Program and Comparison groups. The pre-test scores were used as a covariate to compute the TAKS adjusted post-test means for the experimental and comparison groups. The original means and adjusted means are presented in Table 2. The analyses of covariance results are presented in Table 3 where the pre-test scores were used as a covariate to adjust the original means. The results indicated an $F = 1.02$, $p = .32$, which is not statistically significant. The null hypothesis is therefore retained. The resulting $F$ value indicated that more variance within the groups existed than between the experimental and
comparison group adjusted means. Also, the large sums of squares for the error term suggests that other variables should be included in the analysis of covariance, e.g. gender, teacher, or classroom environment.

Table 2

*Original and Adjusted SES TAKS Social Studies Means*

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Original Mean</th>
<th>Adjusted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>8</td>
<td>2217.38</td>
<td>2214.30</td>
</tr>
<tr>
<td>Comparison</td>
<td>29</td>
<td>2222.93</td>
<td>2223.78</td>
</tr>
</tbody>
</table>

Table 3

*Analysis of Covariance on Low SES TAKS Social Studies Scores*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariate</td>
<td>25017.26</td>
<td>1</td>
<td>25017.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>560.69</td>
<td>1</td>
<td>560.69</td>
<td>1.02</td>
<td>.32</td>
</tr>
<tr>
<td>Error</td>
<td>18692.48</td>
<td>34</td>
<td>549.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>44270.43</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hispanic Differences

H03: The program group Hispanic subpopulation participating in the TEKS/Tools Grant Program (treatment group) will show no statistically significant difference on the social studies 11th grade Exit TAKS test scores as compared to the scores of the comparison group Hispanic subpopulation that did not participate in the TEKS/Tools Grant Program (control group).

The third research question compared TAKS adjusted mean social studies scores from Hispanic subpopulations in the Program and Comparison groups. The pre-test scores were used as a covariate to compute the TAKS adjusted post-test means for the experimental and comparison groups. The original means and adjusted means are presented in Table 4. The analyses of covariance results are presented in Table 5 where the pre-test scores were used as a
covariate to adjust the original means. The results indicated an $F = .50$, $p = .49$, which is not statistically significant. The null hypothesis is therefore retained. It should be noted that the $F$-ratio has an expected value of 1.0 (Hinkle, 1998, 265-266). The resulting $F$ value indicated that more variance within the groups existed than between the experimental and comparison group adjusted means. Once again, the error sums of squares indicates that other variables should be modeled to account for the unexplained variance.

Table 4

*Original and Adjusted Hispanic TAKS Social Studies Means*

<table>
<thead>
<tr>
<th>Group</th>
<th>$N$</th>
<th>Original Mean</th>
<th>Adjusted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>8</td>
<td>2229.88</td>
<td>2219.20</td>
</tr>
<tr>
<td>Comparison</td>
<td>29</td>
<td>2223.48</td>
<td>2226.43</td>
</tr>
</tbody>
</table>

Table 5

*Analysis of Covariance on Hispanic TAKS Social Studies Scores*

<table>
<thead>
<tr>
<th>Source</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariate</td>
<td>40562.60</td>
<td>1</td>
<td>40562.60</td>
<td>.50</td>
<td>.49</td>
</tr>
<tr>
<td>Hispanic</td>
<td>318.34</td>
<td>1</td>
<td>318.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>21821.52</td>
<td>34</td>
<td>641.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>62702.46</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Decatur Sample Trend Analysis

H04: There will be no statistically significant linear trend in the TAKS social study test means across the 9th, 10th, and 11th grades for the Decatur students’ scores only.

The fourth research question was answered using TAKS social studies gain scores from the 9th, 10th, and 11th grades of the students in the Decatur group. The one-way linear trend in the
TAKS social studies gain scores of the Decatur sample was tested using the repeated measures general linear model. The mean and standard deviations for the 9th, 10th, and 11th grades are in Table 6. The repeated measures results are in Table 7. The F = 404.79, p < .0001, indicated a statistically significant linear trend in the Decatur mean gain scores across the three grade levels.

Table 6

_Decatur Sample Mean Gain Scores for 9th, 10th, 11th Grades_

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>9th Grade</td>
<td>122</td>
<td>52.28</td>
<td>16.08</td>
</tr>
<tr>
<td>10th Grad</td>
<td>122</td>
<td>69.43</td>
<td>18.57</td>
</tr>
<tr>
<td>11th Grade</td>
<td>122</td>
<td>80.66</td>
<td>17.09</td>
</tr>
</tbody>
</table>

Table 7

_Repeated Measures for 9th, 10th, 11th Grades (Decatur Sample Only)_

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>49836.76</td>
<td>2</td>
<td>24918.38</td>
<td>404.79</td>
<td>.0001</td>
</tr>
<tr>
<td>Error</td>
<td>14897.24</td>
<td>242</td>
<td>61.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>64734.00</td>
<td>244</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The implications and practical application of the meaning of these findings are discussed in chapter 5. These include recommendations regarding the TEKS/Tools Grant Program strategies and other ramifications from the study. Areas for future research and study are addressed.
CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Schools receiving the TEKS/Tools Grant Program assumed their changes to the structure of the curriculum and instructional delivery would result in some improvement in student scores on the social studies TAKS test (Social Studies Center, 2001). If that assumption were valid, then there would be an improvement in the social studies TAKS test scores of schools participating in the TEKS/Tools Grant Program as compared to the social studies TAKS test scores of the non-participating schools. There would be an improvement in the Hispanic and low SES subpopulations social studies TAKS test scores in the schools participating in the TEKS/Tools Grant Program as compared to the Hispanic and low SES subpopulations social studies TAKS test scores of the non-participating schools. The program target school (Decatur) students would show improvement from grade 9 to 10 to 11 on the social studies TAKS test scores. While the TEKS tools made an impact in a more controlled context of the Decatur schools, lack of consistency of application, implementation and commitment to the program yielded flat results in the Program schools as compared to the comparison group schools. The at-risk subpopulations had little improvement in the program group schools and in the comparison group schools.

The purpose of this study was to learn whether or not any statistically significant differences in social studies scores would exist between the treatment schools (those that participated in the TEKS/Tools Training Program) and the non-treatment group schools (those which did not participate in the TEKS/Tools Training Program). The impact of program implementation was of interest because schools receiving the TEKS/Tools Grant Program
assumed their changes to the structure of the curriculum and instructional delivery would result in some improvement in student scores on the social studies TAKS test (Social Studies Center, 2001). If that assumption were valid, then some improvement would be expected for schools participating in the program. There was also particular interest in the Hispanic and low SES subpopulations which have been shown to be more at risk for low scores and failure on standardized tests. The Decatur sample school scores were subjected to a trend analysis to determine gains over a period of time, from grade 9 to 10 to 11 on the social studies TAKS test scores.

Discussion

The first research question was: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the comparison (control) group that did not participate? The assumption was that the schools that had participated in the intervention would have significantly higher gain scores. I expected that the program group would outperform the control group students who did not participate in the TEKS/Tools Grant Program. However, the findings were opposite of what was expected, that is, the Comparison group had an adjusted mean score that was greater than the program group adjusted mean.

In general, when different teachers implement trainings, different results are to be expected. As an explanation for variance in student improvement scores, Isaac and Michael (1981) stated, “It is dangerous to assume that two teachers are actually teaching the same method. Observers often report critical differences that suggest two versions of a given method. Such an interaction is crucial in making meaningful interpretations” (p. 94). Would more
systematic and rigorous training of teachers be the answer to this problem? Education Secretary Margaret Spellings, an architect of No Child Left Behind seems to think so. In her keynote address at the dedication of the Southern Methodist University School of Education and Human Development (SMU 2005), Spellings highlighted the need for greater consistency in teacher preparation programs necessitated, citing SMU as a potential forerunner in dealing with testing issues and with research and training that align teaching to testing:

Before education reform, we relied too much on good intentions, abstract philosophies and untested, sometimes harmful, fads. If most children were doing well in school, that was considered good enough. It didn't matter if some children were falling behind. And people were in denial—they didn't think the achievement gap was real, or that it mattered. Accountability, assessment and disaggregated data changed all that. Researchers like those at SMU led the way in creating practical tools that teachers could use. (*NY Times*, November 3, 2005, p.19)

The implications of Spelling’s speech are far-ranging. Her assertions regarding the failure of public schools and training colleges to provide teacher preparation programs that effectively prepare teachers to deal with standardized testing reflects a frightening narrowing of the lens regarding the purpose of teacher preparation. As educators, we are all aware of the need for more research-based information regarding ways to prepare teachers effectively to maintain high quality instruction within the high stakes testing environment; however, test preparation cannot become the *raison d’etre* for teacher preparation programs.

One must give Texas credit for supporting staff development efforts in social studies TAKS test preparation. However, the time available for planning and implementing the training and transferring the training into classrooms so that it would make a difference in test scores was limited. The type of top down reform necessitated by such time constraints has its shortcomings. In *Change Forces: Probing the Depths of Educational Reform* (1998), Michael Fullan stresses that the key to successful change is the improvement in relationships between all involved and
not simply the imposition of top down reform. Consistent top-down support, that is, administrative support, was different across schools, as was, teacher motivation and, likely, program implementation across schools within the CCG.

During the verification process, while not formally measured, I noted inconsistencies in training and implementation. The 29 school that made up the non-participating schools had varying levels of preparedness and staffing. In the verification process the contact persons were candid as to their school levels. I noted the narrative accounts by these schools and a sizable number were confidant that they already had in place the strategies and staffing that the TEKS/Tools Grant Program was going to provide. While this was not measured, if teachers were already doing some of the program strategies, this aspect must be considered as a confounding factor in the control groups’ significant gain score performance. In addition, the training took place over several locations (see Appendix B) and the quality of the delivery was unpredictable. The program schools had inherently different and possibly erratic experiences. The classroom application could reflect that uneven training and eventual application of training. An important emphasis in educational change is based on creating the conditions to develop the capacity of both organizations and individuals to learn. The focus moves away from an emphasis on structural change towards changing the culture of classrooms and schools. Therein lie potentially difficult issues for curricular reform, especially in the face of fairly immediate demands in terms of accountability testing.

Governors from 45 states who met in conference early in 2005 acknowledged those very problems in state testing and preparation. Two recommendations from that meeting dealt with aligning programs and boosting the levels and skill knowledge of those directing the schools and their needs (Olsen, 2005). In this study, administrative resistance, teacher quality and quality of
the implementation were not measured and there really was no way to control for differences across the schools as far as implementation.

Several conclusions can be based on these findings: teachers of social studies differ in delivery for the type of teaching that includes a high stakes test and we are not able to measure their effectiveness with existing models. Of math, reading and social studies, the last is the area least attended to as far as academic preparation and therefore it is difficult to assess the true progress of any program change or intervention. We have only a few years of data to look at anywhere in the nation and a long term study is not available at this time. This study of Texas TAKS social studies test scores used the best available resources for comparing the program group and the non program group, yet individual schools may have been at far different stages of preparedness for the tests prior to the provision of the TEKS/Tools Grant Training Program. Indeed, the verification narrative indicates that some non program schools elected not to participate because they believed the necessary program requirements were in place prior to the Grant.

In addition to the limitation of a small sample size in the analysis of gain scores, I must also acknowledge the unbalanced school group sizes used in that analysis. Even though the analysis was weighted, having 9 program schools and 29 comparison schools was problematic in the analysis of student gain scores. Ideally, a group of 20 program schools and 20 comparison schools would have been more powerful. However, the reality of data collection in education is seldom ideal. Future study could benefit from more balanced sample groups. As Texas generates more social studies test performance data, this balance for study may be feasible.

The second research question: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between low socioeconomic
students in the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the low socioeconomic students in the Comparison (control) group that did not participate? Results from the analysis of covariance (ANCOVA) indicated no statistically significant difference between the two groups on adjusted mean scores. Low SES students performed equally well, whether they participated in the TEKS/Tools Grant program or not. The apparently flat improvement for the low SES students is not uncommon.

The CCGs’ poverty statistics placed about 33% of the population at a poverty level. Many school policies and practices benefit middle class and wealthy students at the expense of students in poverty (Darling-Hammond & Post, 2000; Kozol, 1992; Tozer, 2000). M. R. Rank points out that “Americans tend to rank individual reasons (such as laziness, lack of effort, and low ability) as the most important factors related to poverty, while structural reasons such as unemployment and discrimination are typically viewed as less important. It is only logical, then, that most educators, who are primarily middle class, have internalized these attitudes” (Rank 2004 p. 50).

According to the major findings of The Center on Education Policy (2005) High School Exit Exams: Report initial pass rates on tests and achievement gaps have proved to be resistant to change, especially in states where exit exams have been in place for several years. Amrein and Berliner state:

High school graduation exams disproportionately affect students from lower socioeconomic backgrounds. High school graduation exams are more likely to be found in states with the greatest degrees of poverty as compared to the nation. Economically disadvantaged students are most often found in the South and the Southwest and least often found in the Northeast and Midwest. As noted, states in the South and the Southwest are most likely to have high-stakes testing policies. Further, 69% of the states with child poverty levels greater than the nation have or have plans to implement high school graduation exams. Seventy percent of the states with the greatest 1990–1998 increases in the number of children living in poverty have or have plans to implement such exams. Again, if these high-stakes tests are discovered not to have their intended
effects, that is, if they fail to promote transfer of learning and education in its broadest sense, as the nation desires, the mistake will have greater consequences for America's poorest children. (p.10)

Research on students from low socioeconomic backgrounds is flush with factors that impede the achievement of students on high-level thinking, high stakes tests. There is some disagreement and controversy as to why; but there is general agreement that students from low socioeconomic backgrounds are at high-risk low achievement. Major premises in the work of Ruby Payne, a noted expert on children of poverty, include: language issues cause many students from generational poverty not to fully develop the cognitive structures needed to learn at the levels required by state tests; teaching is what happens outside the head; learning is what happens inside the head; for these students to learn, direct teaching must occur to build these cognitive structures; and relationships are the key motivators for learning for students from generational poverty (Payne, 2001). At play in the background of low SES students is the suggestion by Ruby Payne (2001) and Reuven Feuerstein (1980) that students from the generational poverty that make up a large part of Texas low SES students struggle with identity foreclosure and have difficulty seeing themselves as successful and therefore defeat themselves when attempting high stakes assessments and achievement measures that challenge them. This phenomenon needs further study and consideration.

A California researcher, Maria Montano-Harmon (1994), has found that many low-income students do not use formal register and only know casual register. There are five registers of language: frozen, formal, consultative, casual, and intimate. Formal register is standard business and educational language. Formal register features complete sentences and specific word choice. Casual register is characterized by a 400- to 500-word vocabulary, broken sentences, jargon, and many non-verbal assists. State assessments test at the formal register
level. Montano-Harmon and Ruby Payne are among those researchers who believe the achievement gap is due in large part (although they certainly acknowledge other factors) to the language issues. Because the structure and specificity of language is not as readily available to those students who only know casual register, their achievement lags. In addition to language factors, there may be cognitive problems with low SES students that complicate the process. Reuven Feuerstein (1980) and collaborating cognitivists see students from low SES backgrounds in an environment with no routine and structure; this environment impedes the student’s ability to plan. According to Feuerstein (1980): if an individual cannot plan, he or she cannot predict; if an individual cannot predict, he or she cannot identify cause and effect; and if an individual cannot identify cause and effect, he or she cannot identify consequence. The ability to identify cause an effect as well as consequence is a major higher thinking level skill required by the Texas social studies TAKS Exit tests (TEA 2002-2005).

Other researchers stress the inequitable and hostile environments that many children in poverty face at school and in the larger society. Studies show that high-poverty schools implement less rigorous curricula (Barton, 2004), employ fewer experienced teachers (Rank, 2004), have higher student-to-teacher ratios (Barton, 2003; Karoly, 2001), offer lower teacher salaries (Karoly, 2001), have larger class sizes (Barton, 2003), and receive less funding (Carey, 2005; Kozol, 1992) than low-poverty schools. National Commission on Teaching and America’s Future (2004) concludes:

The evidence...proves beyond any shadow of a doubt that children...who come from families with poorer economic backgrounds...are not being given an opportunity to learn that is equal to that offered children from the most privileged families. The obvious cause of this inequality lies in the finding that the most disadvantaged children attend schools that do not have basic facilities and conditions conducive to providing them with a quality education. (p. 7)
The performance of low SES and Hispanic students had much in common. Resolving fundamental questions about the fairness of exit exams and appropriateness of supports for English language learners is crucial if reform is to succeed in helping all students. The 2005 Governor’s Conference, among other recommendations previously cited, called for states to redesign high schools in ways that make them more flexible, smaller, and more attuned to the needs of low-performing students, which the Conference defined as low SES and minority students (Olsen, 2005). Literacy appears to link the two major at-risk groups with 2003 US Department of Education figures indicating that by 12th grade low SES students will be 40% below reading level (as opposed to non low SES who are 24% below reading level) and Hispanic students are 39% below reading level (as opposed to White students who are 21% below level). Research shows that significant achievement gaps in reading continue to exist between White students and poor and/or minority students. For example, in high-poverty schools, more than half of incoming 9th grade students read two to three grade levels behind and, on average, Hispanic 12th grade students read at the same level as White 8th grade students according to US Department of Education 2004 figures. In the discussion of the third research question, which deals with Hispanic students, this common theme of a literacy barrier is also relative to the low SES population in the study.

An aspect of the matching based on the CCG demographic data is notable and may relate to both the low SES and the Hispanic score analysis. I assumed that the formula for match was the best available as referenced by Texas Education Agency. Indeed there may be shortcomings in the demographic aspects of the matching formulas that had confounding effects on the outcomes. The number of low SES and Hispanic populations at the CCG schools was not evenly
matched. The match was best available; however, the schools and TEA acknowledge inconsistencies in reporting.

The third research question: Will there be a statistically significant difference in the mean gain scores on the 11th grade TAKS Social Studies Exit Test between Hispanic students in the Program (experimental) group that participated in the TEKS/Tools Grant Program (treatment) and the Hispanic students in the Comparison (control) group that did not participate? Results from the analysis of covariance indicated that there was no statistically significant adjusted mean score difference between the two groups. Hispanic students performed equally well, whether they participated in the TEKS/Tools Grant program or not.

Reasons for the flat performance of the Hispanic subpopulation in the study relate to prior knowledge and the literacy challenge. Hispanics have not lived America’s history; and, therefore, have little of that story to draw from contextually. As noted by Vygotsky (1978) in his theories on social constructivism, much of what it intimately known is that which is lived. This may also be relevant to English language learners’ (ELLs’) lack of exposure to English. Reading achievement is affected by inadequate prior knowledge and experiences: and the language of academic success is foreign to an ELL (Doty, 2003). In terms of linguistic ability, Hispanic high school students score consistently behind other populations according to studies (NAEP 2004-2005; Padrón, 1994; Cummins 1979&1989).

Differing patterns of literacy exist for and native English speakers in the United States. Research shows that ELLs’ achievement in both reading and writing tends to be lower than that of native English speakers (Padrón, 1994; Truscott & Watts-Taffe, 1998). For example, the reading assessment by the National Assessment of Educational Progress (2002) found dramatic differences between Hispanic students and non-Hispanic white students; 56% of Hispanics in
Grade 4 performed below the basic reading level for their grade, compared to 25% of non-Hispanic white students. Similar differences were found at grades 8 and 12, with Hispanic students reading at significantly lower levels than native-English-speaking students. Such differences place ELLs at risk for school failure, as academic success is highly dependent on literacy skills (Padrón, 1994).

Factors that influence ELLs’ learning of English literacy include their first language literacy, the type of literacy instruction they receive, and uses of literacy in their homes. Research has shown that literacy skills in the first language influence the acquisition of literacy in an additional language. Cummins (1979, 1989) has posited that cognitive academic skills in the first language are likely to transfer to the second language. Recent research supports his view. A study of Spanish-speaking students found that those students who had been explicitly taught to read in Spanish transferred a variety of skills such as phonemic awareness, word reading, word knowledge, and comprehension strategies to English (August, Calderon, & Carlo, 2002). Students with a strong achievement background in their first language are able to utilize metacognitive skills when acquiring literacy skills in another language. For example, students who have developed literacy in their first language may be able to recognize cognates—words that look similar and have the same meaning in both languages. They may also be able to transfer their knowledge of academic strategies such as note taking from their first to their new language. Grabe and Stroller (2002) add that students must have a sufficient amount of knowledge in the second language to make effective use of first language skills to enhance their comprehension in the second language. Average TAKS reading achievement scores for the Hispanic population in 2003 and 2004 lagged behind the White population by 8-10% at all levels
including the 11th grade Exit test (TEA). The comfort level of the Hispanic student testing in academic English is confirmed by the consistent under achievement on the TAKS tests.

Literacy instruction was not an emphasized element of the Social Studies TEKS/Tools Grant Program (SSC). Unfortunately, content literacy strategies training is not universally implemented for Texas social studies teachers. However, it is an area that is recognized by TEA as necessary for the future and the agency has mandated professional development for the fall of 2005 for all teachers in Texas who have English language learners in their classroom (TEA 2005). This initiative may indeed be a reaction to the flat Hispanic progress trend across the state (as well as in this study). The type of literacy instruction that students receive also influences their acquisition of English literacy. Researchers have found that schools that enroll large numbers of students from diverse backgrounds, including ELLs, tend to spend less time on instruction and activities that foster higher-order thinking skills than do schools that serve primarily middle-class, English-speaking students (Au, 2000; Moll & González, 1994). Literacy instruction for ELLs is mostly passive, with little time allocated for active student participation (García, 2000; Truscott & Watts-Taffe, 1998). In a study of eight predominantly urban schools with a majority of Hispanic students, Padrón (1994) found that ELLs remained mostly passive in class, listening to their teachers read to them and not spending much time actively reading themselves. In a study of mainstream classes with ELLs from diverse linguistic backgrounds, Truscott and Watts-Taffe (1998) observed no opportunities for students to read texts at their own reading levels and very little emphasis on reading comprehension instruction. Inherent in this teaching learning disconnect is the problem of no link to the history of the United States or its related symbols including language symbols. Terms that are meaningful to a student who has
grown up with a culture are a difficult concept for those who have not lived the history or its
daily reinforcement (Padrón, 1994).

The third factor that influences literacy instruction is the influence of parental and
community attitudes toward reading and uses of literacy (Grabe & Stoller, 2002). Although
ELLs’ perceptions and uses of literacy at home may differ from those of native English speakers,
this does not mean that literacy activities do not take place in their homes. Research shows that
rich literacy experiences take place in the homes of many ELLs, often in more than one
language. For example, Delgado-Gaitán and Trueba (1991) found that literacy activities in
Hispanic households included children telling stories and singing chants and older siblings
reading to younger ones. Even though Hispanic students’ daily lives outside of school may
remain somewhat of a mystery to their teachers, students’ experiences and knowledge do play a
role in their acquisition of English literacy. Researchers contend that children fare better in
school when their instruction is congruent with their experiences at home (Au & Kawakami,
1994; Gee, 2001; Moll & González, 1994). Teachers need to be aware that although ELLs’
knowledge base may differ from that of native English speakers, they bring rich funds of
knowledge to the classroom (Moll & González, 1994) and additional attention is necessary to
deal with the inherent differences. One way to connect home and school experiences is to
integrate spoken and written narrative forms and to draw narrative content from the culture of the
community. To do this, teachers can have students write about themselves and their culture in
journals. Students can also conduct oral history interviews with each other and with their
families and write biographies based on the interviews. The very study of social studies does
lend itself to that type of enhancement and multicultural perspective. The comprehensive nature
of the TAKS test could be addressed for the Hispanic student with the awareness of the specific
needs of this student. There is a richness of resource for working with this group of students and, as teachers, we must redirect our strategies, include those that connect for the Hispanic student, and create learning structures with an awareness to the needs of this population.

Another linguistic element may be where the TAKS test falls for any given student along his or her developmental continuum between English novice and expert, and how to teach a student whose proficiency is emerging is an ongoing debate. However, the TAKS test for the Hispanic student is a reality and the preparation of that student for the social studies Exit test will depend on strategies that are effective for that student. Jim Cummins’ (1994) research on basic interpersonal communication skills (BICS) (conversation) and cognitive academic language proficiency (CALPS) (academic) emphasizes the difference in levels of learning that resemble the low SES dilemma of casual register language (conversation) versus formal register (academic). The dilemma is those students’ language skills are at a level that is below the level we are testing. Cummins (2001), Padron (1999), Peregoy (2000), and Rivera (1991) agree that from the beginning level of comprehension to the advanced level of comprehension there is a vast range. The five or six (depending on the theorist) levels go from minimal comprehension to some concept detail to the level that allows for abstract reasoning with little difficulty.

The reality is the text and language of social studies is especially difficult. Every text may have a variety of text structures. History texts are typically written in a very descriptive manner, which requires a reader to be able to bring his or her own structure to the text as opposed to when a text has time order, cause-effect structures that are more readily accessible in terms of constructing meaning. Social studies texts use abstract language, long sentences that often include multiple embedded clauses, complex grammar structure, unfamiliar past tense forms, extensive use of pronouns, decontextualized language, and value laden language (Doty,
Couple the language barriers with limited prior knowledge of content and inadequate map and graph reading skills and the task becomes intimidating for the second language learner. Specific strategies for that type of student may differ markedly from the student who is not dealing with those additional problems of language and class (Harvey, 2000).

Again, there is a connection to the problems of testing students from low SES. The relationship of teacher and student is one of the key motivators for learning for students from generational poverty (Payne 2001). Many researchers argue that the relationship between ELLs and their teachers is also central to the development of English literacy. For example, García (1994) posits that effective programs for ELLs show evidence of a highly informal, almost familial social and collaborative relationship between students and teachers. Au (2000) stresses the importance of strong teacher-student relationships that build students’ trust in the teacher and enhance their learning. Moll and González (1994) argue that the relationships between teachers and students influence students’ engagement with the content and skills taught.

Different strategies are needed for students affected by the achievement gap caused by poverty and/or language issues. Social studies testing involves a high literacy level that most students of poverty and those who are not familiar with formal academic English will struggle to achieve. The strategies that work for students not categorized as low SES or for native English speakers (such as repetition, curriculum alignment, concept reinforcement, critical thinking, practice using Bloom’s-based questioning strategies) may be unclear to marginalized populations. Trying to make up for years of missed opportunity and education by the 11th grade diploma-dependent test dates is largely an insurmountable challenge. The conceptual and strategic knowledge base upon which a teacher might build is severely lacking in many students from language different and/or economically disadvantaged backgrounds.
The fourth research question: Is there a statistically significant linear trend in the TAKS social study means across the 9th, 10th, and 11th grades for the Decatur students only? Results from the repeated measures analysis indicated a statistically significant linear trend in the Decatur Sample Group TAKS social studies mean gain scores across the 9th, 10th, and 11th grade levels. The 9th grade mean was 52.28, the 10th grade mean was 69.43, and the 11th grade mean was 80.66. The linear trend or increase across the grade levels may be attributable to a consistent teaching approach and adherence to the TEKS for the Decatur curriculum. Statistical regression can inflate gain scores, especially in instances where lower performing students are exposed to any type of intervention (in spite of the “quality” of the intervention). Issac and Michael discuss this effect in detail (p.98) and emphasize the possibility of this effect on any extreme scores, high or low. Because this type of statistical regression is most likely when pre and posttest scores are correlated less than 1.0, a correlation coefficient was calculated, yielding a result of .765. More importantly, the gain scores continued to improve after the initial year of intervention, providing evidence of real impact by the intervention.

Reasons for the gains made by Decatur can be attributed to several factors. Although we did not measure motivation of teachers and enthusiasm of the administration in the Program or Comparison schools, we did observe a high degree of commitment by both in the Decatur environment. Previous studies have shown that support from the top down is crucial in the success of any reform initiative (Hord 1999) as is “buy in” by the teachers who will implement any program. For the first time in a decade the Decatur staff was empowered to buy into, design, and align the curriculum (SSC, 2002) (DISD, 2004). In writing and implementing the program, the teachers had a high degree of involvement and felt a greater obligation to adhere to the guidelines as they felt ownership of those guidelines. The teachers gained a deeper knowledge
of the entire restructured curriculum by virtue of their participation in the program and the time allowed by that program for individual and team study. The teachers had the opportunity for the first time in the Decatur district to actually align together and learn from each other.

Even though the Decatur sample made significant gains, further evaluation could clarify reasons for those gains. The effect of practice and maturation is of some impact when students test on parallel tests for several years. Decatur students had three times to practice under actual test conditions and educators feel that this has a compounding effect on gains (Lyman, 1991). As students mature they grow wiser and more able to adapt to the testing scenario. There is a strong influence of maturation suggested by developmental theorists in education that traces to the core literature. Piaget (1983) and others make a strong case for the individual developing readiness and ability as maturity evolves.

The Decatur Sample gains in social studies should be viewed with the knowledge of the scores made in other disciplines on the Texas TAKS tests. Over 90% of Decatur students pass social studies on the first try while these same students had pass rates under 70% in science, language arts, and math (DISD, 2004-2005). Administrative and teacher adherence to the program strategies were rigorous and the curriculum was tightly aligned to the test for social studies. The Decatur program participants attribute this large first time test pass rate as a result of that adherence.

Other Considerations and Educational Importance

Other factors that had a possible impact include the short duration of this study. I will continue to collect data to measure future trends. Longitudinal studies are more reliable in showing true achievement gains. The Center on Education Policy (2005) High School Exit
Exams: Report and Northwest Regional Educational Laboratory’s Center for School and District Improvement (2004), both emphasize throughout their reports that data are difficult to analyze and the time frame is short making the measure of any real gain difficult. The small sample size must be acknowledged. It would be desirable to gather more data for a larger sample as that data become available. The age of the sample is a possible factor. It is extremely difficult to make up for lost time in a well established program and we have in this situation a program that is innovative but quite new relative to other areas of testing. The Center on Education Policy (2005) High School Exit Exams: Report indicated that gains are being made at lower levels of testing accountability and with time, perhaps, that improvement will carry through a student’s high school testing years.

The conclusions for this study support the continued evaluation of the effects of the type of intervention employed with the TEKS/Tools Grant Program for Decatur. The intervention strategies employed in the program are considered best practices and the alignment of the TEKS and the TAKS test seem to indicate a positive effect on the instructional gains for students. The expected difference in subpopulations (SES and Hispanic) of interest was not indicated although research and practice suggest an impact when dealing with at risk populations.

One consideration for the lack of specific data showing significant gains by the treatment groups in the study must be noted. The TEKS/Tools Grant program contains some elements that by nature merit a longer period of study. Two or three years does not seem an adequate time frame to evaluate changes in curricular approach, design and delivery that start when a student is in the elementary grades and build until that student is in high school. A cumulative curriculum, as the TEKS are by nature, would, therefore, require a longer period of study to determine whether that approach produces a significant impact in student knowledge and improvement on
the TAKS exam. The educational findings of this study can be summarized as all students, whether in the experimental or control group, showed consistent improvement from 10th to 11th grade on the TAKS social studies test.

With great emphasis, I must conclude that the weight of a high stakes test in a subject area that is not viewed with parallel importance and priority by all stakeholders in the federal or state systems that make the policies must be questioned. Consistently the tenuous nature of the position of social studies in the testing and accountability movement and the murky view of where, when, and how social studies should be tested gives no clear compass for a program, administrator, teacher, or student to follow.

Recommendations

Directions for Future Research

Based on the findings and conclusions of this study, the following recommendations are offered:

1. A more comprehensive study should be conducted to increase the sample size. Additional years of TAKS scores and additional districts would increase the power of the study.

2. A more comprehensive study using individual student scores rather than a mean score for each district would be desirable and more precise. (The Decatur sample method for additional districts would increase the number and power of the study).

3. A correlational study of social studies TAKS scores and reading scores and social studies TAKS scores and math scores could be helpful to educators for remediation, planning, and placement of students. With the implications for success on the social studies TAKS and the inherent link to literacy, the relationship of reading ability is critical. The inclusion of more maps and graphs than TAAS or End of Course tests contain may present barriers requiring interpretation by students whose math skills are not practiced in related map or graph skills.

4. Are all the students identified as gifted achieving mastery as evidenced by a commended score? Based on preliminary data further study of the performance of identified gifted
students would be helpful to determine whether schools are meeting that subpopulation’s needs.

5. Are the low SES and Hispanic students who are identified as gifted achieving commended scores at the same rate as their non-low SES and White counterparts? If teachers are missing the mark with these subpopulations as a group, then are the most gifted among these sub populations being served?

7. Preliminary examination of the data indicates possible ceiling effects on the TAKS social studies test. Further study is recommended to determine the existence and/or extent of this phenomenon.

8. Preliminary examination of the data indicated a countertrend in gender performance. Further study on the relationship of gender to the test is indicated by the gender gap that was not seen as a pattern for previous state social studies tests using the score history for EOC and TAAS from 1996 to 2000 as a guide.

9. A study of teachers’ efficacy in teaching the TEKS, namely, do they align TEKS with TAKS.

10. Further study of specific strategies to enhance the performance of students at the lower levels of language proficiency (which includes low SES and Hispanic subpopulations).
APPENDIX A

SOCIAL STUDIES HIGH STAKES TESTING BY STATE
<table>
<thead>
<tr>
<th>State</th>
<th>Name of Test</th>
<th>Grade Admin</th>
<th>Type of Test</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>Alabama High School Graduation Exam</td>
<td>11</td>
<td>standards based</td>
<td>2004</td>
</tr>
<tr>
<td>Georgia</td>
<td>Georgia High School Graduation Test</td>
<td>11</td>
<td>standards based</td>
<td>1996</td>
</tr>
<tr>
<td>Louisiana</td>
<td>Graduation Exit Exam for the 21st Century</td>
<td>10 &amp; 11</td>
<td>standards based</td>
<td>2001</td>
</tr>
<tr>
<td>Maryland</td>
<td>Maryland High School Assessments (Govt. Only)</td>
<td>varies</td>
<td>end of course</td>
<td>2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2008 for diploma)</td>
</tr>
<tr>
<td>Mississippi</td>
<td>Mississippi Subject Area Testing Program</td>
<td>varies</td>
<td>end of course</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>(US History since 1877 only)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>New Mexico High School Competency Exam</td>
<td>10</td>
<td>minimum competency</td>
<td>1990</td>
</tr>
<tr>
<td>New York</td>
<td>Regents Comprehensive Examinations</td>
<td>varies</td>
<td>end of course</td>
<td>1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2000 for diploma)</td>
</tr>
<tr>
<td>Texas</td>
<td>Texas Assessment of Knowledge &amp; Skills</td>
<td>11</td>
<td>standards based</td>
<td>2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2005 for diploma)</td>
</tr>
<tr>
<td>Virginia</td>
<td>Standards of Learning End of Course Exams</td>
<td>varies</td>
<td>end of course</td>
<td>1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2004 for diploma)</td>
</tr>
</tbody>
</table>

*from State High School Exams Put to the Test, written by Keith Gayler, Naomi Chudowsky, Nancy Kober, and Madeline Hamilton as a report by the Center on Education Policy, Washington, D.C. August 2003*
APPENDIX B

PRELIMINARY TEKS/TOOLS GRANT INFORMATION
Information Pertaining to Statewide Assessment in Social Studies

In 1999, the 76th Session of the Texas Legislature passed Senate Bill 103. The bill affects the social studies field in several ways. It mandated that the agency develop and implement assessment instruments for Grades 3 through 11, including social studies at Grades 8, 10, and 11 (exit level). All assessment instruments will eventually become a part of the agency’s accountability system.

TEA is currently involved in developing objectives for the new testing program (TAAS II). In October 2000, the agency distributed the second of two surveys. Administrators and teachers statewide reviewed and commented upon proposed Grade 8, 10, and 11 social studies TAAS II objectives and Texas Essential Knowledge and Skills (TEKS) student expectations. TEA received 1,424 social studies response forms. In February 2001, two separate teacher committees met to review the data gleaned from the surveys. A national panel of educational experts subsequently reviewed the data and made recommendations to the agency. TEA is considering all data and feedback and will release the final TAAS II objectives and student expectations in Spring 2001. The agency continually strives to provide Texas teachers with the most current information. The TEA website, www.tea.state.tx.us, is a primary avenue by which teachers and administrators can remain informed.

Statewide Assessments in Social Studies and the Agency’s Accountability System

In July 2000, Commissioner of Education Jim Nelson announced the final accountability rating standards for 2001 and 2002. At the same time, he released a tentative blueprint for 2003 through 2005. Beginning in 2002, the Grade 8 Social Studies TAAS test will become a part of the accountability system for the first time. The agency will not issue accountability ratings in 2003. The new examinations at all Grades (3 through 11) are expected to be more rigorous than
the TAAS tests, due to both the increased requirements for exit-level and the assessment of the TEKS curriculum. Given the significant redesign of the assessment system scheduled for 2003, the blueprint for 2003 through 2005 is subject to change. The complete text of the commissioner’s letter appears at: www.tea.state.tx.us/taa/perf000731.html.

Social Studies Center for Educator Development (SSC)

TEA has awarded a $500,000 grant to Education Service Center Region VI, in collaboration with Texas A&M University and Sam Houston State University, to create a social studies center for professional development in curriculum and assessment. The primary purpose of the center is to assist social studies teachers and administrators statewide in their efforts to prepare students for the new assessments mandated by Senate Bill 103. The center is fortunate to have Karen Wiggins as its full-time director. Many of you know Karen and her outstanding work as the social studies supervisor for Richardson ISD. Under Karen’s leadership, the center is charged with meeting the following goals:

• to assist in on-going efforts to implement the Texas Essential Knowledge and Skills (TEKS) for Social Studies
• to develop effective products and deliverables designed to assist districts statewide in preparing students for new social studies assessments at Grades 8, 10, and 11 exit level
• to exemplify the principles that undergird all social studies programs of excellence

Social Studies Center for Educator Development (SSC) Assistance

Staff members of the new Social Studies Center (SSC) have been hard at work creating the professional development segments, known as TEKS Tools, to present in workshops during the spring and summer and in web-based training for educators who cannot attend the workshops. A key element of this statewide training initiative is that all participants in these
“train-the-trainers” workshops will be expected to replicate the training upon returning to their campus, district, and/or service center region. Social Studies Center staff will conduct a two-day TEKS Tools workshop in Austin for Education Service Center representatives and social studies supervisors on April 3 and 4 at the Austin North Hilton (near intersection of Interstate 35 and Highway 290). During the month of June, three-day regional workshops will be held around the state for teams of teachers (elementary, middle, & high school) and administrators (campus & district levels). The regional workshops are scheduled:

<table>
<thead>
<tr>
<th>Date Range</th>
<th>Area</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 5th through 7th</td>
<td>Houston Area</td>
<td>Fort Bend ISD</td>
</tr>
<tr>
<td>June 13th through 15th</td>
<td>West Area</td>
<td>ESC Region 18 in Midland</td>
</tr>
<tr>
<td>June 18th through 20th</td>
<td>South &amp; Central Areas</td>
<td>ESC Region 1 in Edinburg</td>
</tr>
<tr>
<td>June 25th through 27th</td>
<td>Metroplex Area</td>
<td>Hurst-Euless-Bedford ISD</td>
</tr>
</tbody>
</table>

The TEKS Tools training segments for the workshops and web-based training include:

- Data Driven Decision-Making
- Elementary Social Studies: Building a Foundation
- United States History: Developing a Sense of Time and Place
- World History and World Geography: Connecting Concepts
- Economics, Government, and Culture: 21st Century Citizenship
- Integrating Content and Skills Instruction

The web-based training will be available later in the summer. Information regarding registering for the two-day April workshop or the three-day June workshops will be sent to Education Service Center representatives for distribution to school districts. In addition to training, the Social Studies Center’s website provides many wonderful resources to assist teachers and administrators in their efforts to implement the TEKS. The new website address is:
http://socialstudies.tea.state.tx.us. Explore the site to find the TEKS with glossaries and biographies by grade level and course and annotated bibliographies for each grade level and course. Other resources that may be downloaded include the Texas Social Studies Framework, Social Studies Tool Kit 2000, CD-ROM reviews, Aligning TEKS with PDAS, and Resource Evaluation Form. There is also a listing of state and national organizations related to social studies as well as information and websites for a number of state and national museums. Information on the web site will be updated periodically, so check the site on a regular basis. For additional information on Social Studies Center activities and products, contact the Center’s project director, Karen Wiggins, at the following e-mail address: kwiggins@airmail.net
APPENDIX C

HOW CAMPUS COMPARISON GROUPS FORMED FOR 2002-2003 YEAR
Overview

The comparable improvement (CI) measures depend on campus comparison groups. Each campus has a unique comparison group of 40 other campuses in the state that closely match the target school on a number of characteristics. Comparison groups are recreated each year to account for changes in demographics that may occur. They are used for all group statistics reported on campus AEIS reports and the School Report Cards.

CI in the Texas public school accountability system is computed for TAAS reading and mathematics only, using students tested in English who can be matched by their student identification information to their results from a prior school year; is a campus measure only; and is used for the Gold Performance Acknowledgment system, TSSAS awards (when funded), AEIS reports, and School Report Cards.

The 2002 CI is computed only for schools rated in August 2002 and evaluated under the standard accountability procedures described in this manual. Schools evaluated under the optional alternative accountability procedures in August, even if they later receive a 2002 rating under standard procedures, will have no CI comparison group or measures determined. Results for the entire state are used in the CI calculations; to add a school at a later date has the potential to change CI results for other campuses.

Background

CI has been a statutory component of the accountability system since it began in 1993. Implementation of CI was postponed until the 1995-96 school year when student-level TAAS growth measures became available. Although the Texas Education Code defines the structure of the Texas public school accountability system, it delegates the operational decisions of applying such a system to the Commissioner of Education. Since the specifics of its definition and its
application to the system are not codified, both the calculation and application of Comparable
Improvement are the Commissioner's responsibilities. Texas Education Code §39.051(c) defines
CI and is reprinted in Section XV, Appendix A.

Building Campus Comparison Groups: Characteristics Used

The characteristics used to construct the campus comparison groups include those
defined in statute as well as others found to be statistically significant. The six campus-level
characteristics used in 2001-02 are:

percent of 2001-02 students identified as African American;
percent of 2001-02 students identified as Hispanic;
percent of 2001-02 students identified as White;
percent of 2001-02 students identified as Economically Disadvantaged;
percent of 2001-02 students identified as Limited English Proficient (LEP); and
percent of mobile students as determined from 2000-01 cumulative attendance.

The characteristics analyzed to construct the comparison groups are defined below. Both special
education and non-special education students are counted in the calculations, which are rounded
to one decimal place.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Calculation</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent African American</td>
<td>Number of African American Students Enrolled X 100</td>
<td>2001-02 PEIMS Submission 1</td>
</tr>
<tr>
<td></td>
<td>Students in Campus Membership</td>
<td></td>
</tr>
<tr>
<td>Percent Hispanic</td>
<td>Number of Hispanic Students Enrolled X 100</td>
<td>2001-02 PEIMS Submission 1</td>
</tr>
<tr>
<td></td>
<td>Students in Campus Membership</td>
<td></td>
</tr>
<tr>
<td>Percent White</td>
<td>Number of White Students Enrolled X 100</td>
<td>2001-02 PEIMS Submission 1</td>
</tr>
<tr>
<td></td>
<td>Students in Campus Membership</td>
<td></td>
</tr>
</tbody>
</table>
Percent Economically Disadvantaged
Number of Economically Disadvantaged Students Enrolled X 100
Students in Campus Membership

2001-02 PEIMS Submission 1

Percent LEP
Number of Limited English Proficient Students Enrolled X 100
Students in Campus Membership

2001-02 PEIMS Submission 1

Percent Mobile
Students in Campus Membership* less than 83% of Days Taught X 100
Students in Campus Membership

2000-01 PEIMS Submission 3

How Groups Are Constructed

A unique comparison group of 40 campuses is identified for each school. The group is selected based on the most dominant characteristics of the target campus. The order of dominance is determined by ranking the characteristics from highest to lowest percent. Only schools of similar type (elementary, middle, high school, or multi-level) form the selection pool. Based on the most dominant characteristic for the target school - the one with the highest percentage - from the six listed above, the 100 most similar campuses are selected. That group is further refined by the next most dominant feature, and so on, until 50 comparison campuses are identified. Finally, 10 campuses with the most dissimilar of the less predominant characteristics are eliminated to bring the group size to 40. Only the accountability student group characteristics - African American, Hispanic, White, and Economically Disadvantaged - are used for this final reduction from 50 to 40 campuses; the percent LEP and percent mobile students are not considered when eliminating the least predominant characteristics in this final step.

Example:
Elementary
Campus X: 19.8% Hispanic, 50.3% African American, 29.9% White, 40.4% Economically Disadvantaged, 12.0% LEP, 15.2% Mobile
Step 1: 100 elementary campuses having percentages closest (both above and below) to 50.3% African American students are identified.

Step 2: 10 schools from the initial group of 100 are eliminated on the basis of being most distant from the value of 40.4% Economically Disadvantaged.

Step 3: 10 of the remaining 90 schools that are most distant from 29.9% White students are eliminated.

Step 4: 10 of the remaining 80 schools that are most distant from 19.8% Hispanic students are eliminated.

Step 5: 10 of the remaining 70 schools that are most distant from 15.2% Mobile students are eliminated.

Step 6: 10 of the remaining 60 schools that are most distant from 12.0% LEP students are eliminated.

Step 7: 10 of the remaining 50 schools that are most distant from 29.9% White students and / or 19.8% Hispanic students are eliminated.

The final group size is 40 schools. There is no limit to the number of comparison groups of which a school may be a member. It is theoretically possible for a school to be a member of no comparison group other than its own, or all of them within a particular school type (e.g., high school).

Performance Measured for Comparable Improvement

According to statute, Comparable Improvement must be calculated for assessment results only, specifically for the TAAS. Comparable Improvement measures are based on analysis of growth on the Texas Learning Index (TLI), derived from the English language TAAS reading and mathematics tests that are given at grades 3 - 8, and 10. There are no TLI values available for the Spanish versions of TAAS that are administered in grades 3 - 6, or for TAAS writing.

Campus Comparable Improvement is not based on a single measure; there are multiple calculations involved to determine all components of the Comparable Improvement report and the Gold Performance Acknowledgment criteria. These measures are described in detail.

Identifying Matched Students
Growth measures based on the TLI in reading and mathematics are determined for those students who took the test(s) in the current and prior years. The method for identifying matched students in 2002 is described in detail below. The grade level of the student does not limit the matching in the prior year; retained as well as promoted students can be part of the set of matched students.

**Grades 4 - 8**

Students (non-special education and special education) tested in English in the spring 2002 TAAS administrations on reading and/or mathematics who: are in grades 4, 5, 6, 7, or 8; are part of the 2002 accountability subset (tested non-special education and special education students who were enrolled in the district as of October 26, 2001); and can be matched back to the spring 2001 TAAS administration (English language) in grades 3, 4, 5, 6, 7, or 8, all students, anywhere in the state. NOTE: Although these students must be in the accountability subset in 2002, they do not have to be in the 2001 accountability subset to be matched for Comparable Improvement. Both special education and non-special education TAAS data files are searched to locate prior year results.

**Grade 10**

Students (non-special education and special education) tested in English in the spring 2002 TAAS administrations on reading and/or mathematics who: are in grade 10; are part of the 2002 accountability subset; and can be matched back to the spring 2001 grade 10 TAAS administration, or the spring 2000 or spring 1999 grade 8 TAAS administrations, all students, anywhere in the state.

**Grade 3**
Students tested in the spring 2002 TAAS administrations on reading and mathematics in grade 3 cannot contribute to 2002 Comparable Improvement. NOTE: Campuses without TAAS results at grades 4 - 8 or 10 are paired in order to calculate Comparable Improvement. The exception is campuses serving grades prekindergarten and / or kindergarten only; those schools are not required to be rated in the accountability system. (Refer to Section VI, Special Issues and Exceptions for details on pairing.)

Spanish TAAS

Students taking the Spanish version of the TAAS in grades 3 - 6 cannot contribute to the computation of Comparable Improvement. As there is no TLI measure for Spanish TAAS (because there is no exit-level Spanish TAAS with which to align the results), neither growth nor high-performing status can be determined for these examinees on a comparable scale to those students tested on English versions of the TAAS.

Growth on the Texas Learning Index

Comparable Improvement measures are based on analysis of TLI growth for all matched students in reading and mathematics. The measures take several steps to compute; the process begins with student-level calculations that are then aggregated to the campus level, and those results finally are analyzed within the comparison group. A Texas Learning Index score is preceded by a digit representing the grade tested - 3, 4, 5, 6, 7, 8, or X (exit-level). For example, a student with a TLI mathematics score of 4-78 earned a TLI score of 78 on the 4th grade mathematics TAAS. The top and bottom end of the score range may differ from subject to subject, depending on how much easier or harder the test is at any particular administration. Within a subject, TLIs can be compared to determine the growth between test administrations.
Matched students for reading and matched students for mathematics are separately identified, and individual growth is determined by subject. TLI growth calculations performed at the student level are illustrated below:

TLI Growth (Mathematics) = Current Year Mathematics TLI minus Prior Year Mathematics TLI
TLI Growth (Reading) = Current Year Reading TLI minus Prior Year Reading TLI

A TLI growth of zero means that one year's growth has occurred. A negative value means that less than one year's growth has occurred and a positive value means that more than one year's growth has occurred. Examples of the reading calculation for two sixth grade students are provided:

Example: Jill
   (6-65) - (5-55) = (+10)
   Jill’s performance in Spring 2002 - Jill’s performance in Spring 2001
Example: Jack
   (6-75) - (5-80) = (-5)

Determining the Comparable Improvement Quartile

Step 1: Students Included

Matched students tested on TAAS in 2002 that scored less than a TLI of 85 on the prior year TAAS administration are included in the calculations of TLI growth that determine the Comparable Improvement quartile. (At exit-level, the prior year is either 2001, 2000, or 1999; for all other grades, the prior year is 2001.) These students are the basis of the calculations described in Step 2 through Step 3.

Step 2: Campus Average TLI Growth

For each subject, the student TLI growth values are aggregated to the campus level to create a TLI Average Growth (TAG) for each campus. The calculations, rounded to two decimal places, are illustrated below:
TAG (Reading) = \frac{\text{Sum of Matched Student TLI Growth Values for Reading}}{\text{Total Number of Matched Students in Reading}}

TAG (Mathematics) = \frac{\text{Sum of Matched Student TLI Growth Values for Mathematics}}{\text{Total Number of Matched Students in Mathematics}}

Step 3: Quartile Distribution of Growth

Within the comparison group, TAG values are ranked to determine the quartiles. Each campus is separately assigned one of the following quartile values for reading and for mathematics: Q1 (top 25 percent); Q2 (in the top 50 percent, but not in the top 25 percent); Q3 (in the bottom 50 percent, but not in the lowest 25 percent); Q4 (lowest 25 percent).

Since campuses have a comparison group of 40 schools, usually 10 will comprise each quartile. For each subject, those in Q1 are the 10 schools with the highest TAG within the group; those in Q4 are the 10 schools with the lowest TAG. The number of schools in each quartile can differ if TAG values are tied near the quartile separation points, or if some schools are excluded because they do not meet minimum size criteria. (See below.) Each school is assigned two quartile values - one for reading and one for mathematics - depending on where the TAG falls in the distribution of its unique set of 40 comparison schools. These are the Comparable Improvement measures for the target campus. The quartile value of any school in a comparison group is appropriate only for that comparison group. A school that is a member of multiple comparison groups could have different quartile values for the same performance because that determination depends on the performance of the other schools in each group. The quartile values for a particular target school define Comparable Improvement for the target school only.
MINIMUM SIZE CRITERIA FOR COMPARABLE IMPROVEMENT: A campus (either the target campus or a member of the comparison group) must have at least 10 matched students in a subject to receive a quartile value for that subject. The target campus is not assigned a quartile value for a subject when fewer than 24 campuses in the comparison group meet the minimum matched students criteria.

Other Measures

The AEIS campus Comparable Improvement report provides other TLI-based measures for each target campus and its comparison group of 40 schools. These are detailed below.

TLI Average Growth (TAG) for Prior Year Failers

For both mathematics and reading, the average TLI growth for matched students who failed the TAAS in the prior year is calculated and reported. On the CI report, this is column (4). (See sample CI pages.) The calculations, rounded to two decimal places, are illustrated below:

\[
\text{Prior Year Failer Average TLI Growth (Mathematics)} = \frac{\text{For Matched Students who Failed Mathematics in the Prior Year, Sum of TLI Growth Values for Mathematics}}{\text{Total Number of Matched Students in Mathematics who Failed in the Prior Year}}
\]

\[
\text{Prior Year Failer Average TLI Growth (Reading)} = \frac{\text{For Matched Students who Failed Reading in the Prior Year, Sum of TLI Growth Values for Reading}}{\text{Total Number of Matched Students in Reading who Failed in the Prior Year}}
\]

High-Performing Students

For both mathematics and reading, the percent of matched students excluded from the CI growth analysis because they had a TLI at or above 85 in the prior year is calculated and reported for each target campus and its comparison group of 40 schools. On the CI report, this is column (5). This measure is a component of the Additional Acknowledgment standards for campus
Comparable Improvement. See Section IV, 2002 Acknowledgments and Recognitions for details on the standards and criteria for acknowledgment. These measures, based on total matched students, indicate the percent of total matched students who were excluded from the campus TAG calculations and Comparable Improvement analysis.

The calculations, rounded to one decimal place, are illustrated below:

\[
\text{Percent of High Performing Students (Mathematics)} = \frac{\text{Number of Matched Students with a Prior Year TLI value } \geq 85.0 \text{ for Mathematics}}{\text{Total Number of Matched Students in Mathematics}}
\]

\[
\text{Percent of High Performing Students (Reading)} = \frac{\text{Number of Matched Students with TLI Growth Values } \geq 85.0 \text{ for Reading}}{\text{Total Number of Matched Students in Reading}}
\]

**Rationale for Exclusions of High- or Low-Performing Students**

The Texas Learning Index upon which Comparable Improvement is based is least sensitive to exceptionally high or low performance. This is a direct consequence of the criterion-based design of the state assessment program. Criterion-referenced tests are constructed to determine an individual's level of performance on specific content. In contrast, norm-referenced tests are designed to determine where a person stands relative to a population of examinees on the content being tested. Because criterion-referenced tests are not designed to measure the full extent of one's skills or knowledge, there are inherent "floors" and "ceilings" in the scores one can obtain. Therefore, growth measured when overall performance is exceptionally high or low are likely not very reliable indicators of either performance problems or improvement.

HIGH PERFORMERS. Growth for students scoring a TLI of 85 or above in the first year of the comparison is difficult to measure; therefore the performance of these students is excluded from the calculation of campus average TAGs. Statewide, average TLI growth between 2000 and
2001 tended to be negative when the prior year score was near the top of the scale. The average mathematics TLI growth values turn negative when the previous year's TLI score is 85 or higher, while reading growth turns negative at a prior year score of 91. For consistency, the prior year score of 85 was selected as the threshold for both subjects. Exclusion of these high performers tends to increase the average TLI growth for individual campuses. Between 2000 and 2001, statewide TLI growth in reading was 5.91 without the high performers, but only 1.81 if they were included. For mathematics, statewide TLI growth was 4.03 without the high performers, and 1.26 with the high performers.

LOW PERFORMERS. To address the measurement problems of the lowest performing students, matched students receiving the minimum possible score in either year are excluded from the CI analysis. This action affects very few students; in 2001, 352 mathematics and 492 reading scores out of 1.3 million scores in each subject were excluded.

2002 Comparable Improvement Reports

In September 2002, Comparable Improvement reports are accessible through the Texas Education Agency's website. A CI report is included with the campus AEIS reports transmitted in the fall. The report includes two pages of information: the demographic characteristics used to determine the comparison groups and the TLI growth measures. Annotated samples of these reports are included for illustration on pages 61-62. Acknowledgment on Comparable Improvement in Reading and Mathematics will be printed: on campus data tables provided with the GPA release; on the cover page of the campus AEIS reports; and on page 2 of the corresponding CI report.

All of the data supporting the acknowledgment are printed on the campus CI report, published in October. However, these reports will be accessible from the TEA website in
September, concurrent with the 2002 GPA release, through a link from the campus GPA data table.

APPENDIX D

TEST DEVELOPMENT PROCESS
Texas educators - classroom teachers, curriculum specialists, administrators, and education service center staff - play a vital role in all phases of the test development process. Thousands of Texas educators have served on one or more of the educator committees involved in the development of the state assessment program. These committees represent the state geographically, ethnically, by gender, and by type and size of school district. The procedures described below outline the process used to develop a framework for the tests and provide for the ongoing development of test items.

1) Committees of Texas educators review the state-mandated curriculum to develop appropriate assessment objectives for a specific grade and/or subject test. For each subject area, educators provide advice on an assessment model or structure that aligns with good classroom instruction.

2) Educator committees work with the Texas Education Agency (TEA) both to prepare draft test objectives and to determine how these objectives would best be assessed. These preliminary recommendations are distributed widely for review by teachers, curriculum specialists, assessment specialists, and administrators.

3) A draft of the objectives and student expectations to be assessed is refined based on input from Texas educators. TEA begins a statewide opportunity-to-learn study.

4) Prototype test items are written to measure each objective and, when necessary, are piloted by Texas students from volunteer classrooms.

5) Educator committees assist in developing guidelines for assessing each objective. These guidelines outline the eligible test content and test-item formats and include sample items.

6) With educator input, a preliminary test blueprint is developed that sets the length of the test and the number of test items measuring each objective.
*7) Professional item writers, many of whom are former or current Texas teachers, develop items based on the objectives and the item guidelines.

*8) TEA curriculum and assessment specialists review and revise the proposed test items.

*9) Item-review committees composed of Texas educators review the revised items to judge the appropriateness of item content and difficulty and to eliminate potential bias.

*10) Items are revised again based on input from Texas educator committee meetings and are field-tested with large representative samples of Texas students.

*11) Field-test data are analyzed for reliability, validity, and possible bias.

*12) Data-review committees composed of Texas educators are trained in statistical analysis of field-test data and review each item and its associated data. The committees determine whether items are appropriate for inclusion in the bank of items from which test forms are built.

13) A final blueprint that establishes the length of the test and the number of test items measuring each objective is developed.

*14) All field-test items and data are entered into a computerized item bank. Tests are built from the item bank and are designed to be equivalent in difficulty from one administration to the next.

*15) Tests are administered to Texas students; and results are reported at the student, campus, district, regional, and state levels.

*16) Stringent quality control measures are applied to all stages of printing, scanning, scoring, and reporting.

*17) All Texas Assessment of Knowledge and Skills (TAKS) and Reading Proficiency Tests in English (RPTE) tests are released to the public at the end of each school year.
18) The State Board of Education uses impact data, the statewide opportunity-to-learn study along with additional information to set a passing standard for each new test.

*19) A technical digest that provides verified technical information about the tests to schools and the public is developed annually.

* These steps are repeated annually to ensure that tests of the highest quality are developed.

Further information about the state assessment program is available on the TEA web site: (www.tea.state.tx.us/student.assessment).
APPENDIX E

GAIN SCORES - CCG SCHOOLS
<table>
<thead>
<tr>
<th>District</th>
<th>scaled mean score</th>
<th>scaled mean score</th>
<th>gain score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smithville</td>
<td>2224</td>
<td>2280</td>
<td>56.00</td>
</tr>
<tr>
<td>Marble Falls</td>
<td>2215</td>
<td>2276</td>
<td>61.00</td>
</tr>
<tr>
<td>Farmersville</td>
<td>2193</td>
<td>2263</td>
<td>70.00</td>
</tr>
<tr>
<td>Columbus</td>
<td>2192</td>
<td>2291</td>
<td>99.00</td>
</tr>
<tr>
<td>Palmer</td>
<td>2166</td>
<td>2221</td>
<td>55.00</td>
</tr>
<tr>
<td>Red oak</td>
<td>2204</td>
<td>2266</td>
<td>62.00</td>
</tr>
<tr>
<td>Klein Oak</td>
<td>2253</td>
<td>2286</td>
<td>33.00</td>
</tr>
<tr>
<td>Klein Collins</td>
<td>2283</td>
<td>2310</td>
<td>27.00</td>
</tr>
<tr>
<td>Katy</td>
<td>2283</td>
<td>2334</td>
<td>51.00</td>
</tr>
<tr>
<td>Maypearl</td>
<td>2187</td>
<td>2267</td>
<td>80.00</td>
</tr>
<tr>
<td>Pampa</td>
<td>2137</td>
<td>2272</td>
<td>135.00</td>
</tr>
<tr>
<td>Pinetree</td>
<td>2246</td>
<td>2289</td>
<td>43.00</td>
</tr>
<tr>
<td>Sulphur Springs</td>
<td>2202</td>
<td>2233</td>
<td>31.00</td>
</tr>
<tr>
<td>Alvarado</td>
<td>2156</td>
<td>2220</td>
<td>64.00</td>
</tr>
<tr>
<td>Lampasas</td>
<td>2240</td>
<td>2289</td>
<td>49.00</td>
</tr>
<tr>
<td>Dayton</td>
<td>2179</td>
<td>2230</td>
<td>51.00</td>
</tr>
<tr>
<td>Groesbeck</td>
<td>2145</td>
<td>2232</td>
<td>87.00</td>
</tr>
<tr>
<td>LubCooper</td>
<td>2238</td>
<td>2270</td>
<td>32.00</td>
</tr>
<tr>
<td>LaGrange</td>
<td>2242</td>
<td>2271</td>
<td>29.00</td>
</tr>
<tr>
<td>Clearcreek</td>
<td>2262</td>
<td>2318</td>
<td>56.00</td>
</tr>
<tr>
<td>Shallowater</td>
<td>2314</td>
<td>2343</td>
<td>29.00</td>
</tr>
<tr>
<td>Greenwood</td>
<td>2224</td>
<td>2272</td>
<td>48.00</td>
</tr>
<tr>
<td>Willis</td>
<td>2195</td>
<td>2272</td>
<td>77.00</td>
</tr>
<tr>
<td>Breckenridge</td>
<td>2154</td>
<td>2254</td>
<td>100.00</td>
</tr>
<tr>
<td>Boswell</td>
<td>2195</td>
<td>2281</td>
<td>86.00</td>
</tr>
<tr>
<td>Grape creek</td>
<td>2229</td>
<td>2268</td>
<td>39.00</td>
</tr>
<tr>
<td>Georgetown</td>
<td>2245</td>
<td>2294</td>
<td>49.00</td>
</tr>
<tr>
<td>Hutto</td>
<td>2243</td>
<td>2280</td>
<td>37.00</td>
</tr>
<tr>
<td>Westwood</td>
<td>2392</td>
<td>2418</td>
<td>26.00</td>
</tr>
<tr>
<td><strong>Treatment Group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leander</td>
<td>2238</td>
<td>2265</td>
<td>27.00</td>
</tr>
<tr>
<td>A&amp;M Con</td>
<td>2309</td>
<td>2342</td>
<td>33.00</td>
</tr>
<tr>
<td>Bell</td>
<td>2255</td>
<td>2301</td>
<td>46.00</td>
</tr>
<tr>
<td>Decatur</td>
<td>2184</td>
<td>2260</td>
<td>76.00</td>
</tr>
<tr>
<td>Martin</td>
<td>2277</td>
<td>2323</td>
<td>46.00</td>
</tr>
<tr>
<td>Pearce</td>
<td>2414</td>
<td>2397</td>
<td>-17.00</td>
</tr>
<tr>
<td>Belton</td>
<td>2237</td>
<td>2284</td>
<td>47.00</td>
</tr>
<tr>
<td>Marion</td>
<td>2188</td>
<td>2220</td>
<td>32.00</td>
</tr>
</tbody>
</table>
REFERENCES


Carnoy, M., Loeb, S. & Smith, T.L. (2000, April). Do higher test scores in Texas make for better


Crocco, Margaret. (June 2004). Author interview. Teachers College at Columbia University. New York, NY.


Simon & Schuster.


Montano-Harmon, Maria Rosario. (1994). Presentation given to Harris County Texas, Department of Education on the topic of her research findings.


Herszenhorn, David. (June 7, 2005). Test scores used to analyze school roles. New York Times


Kappan opinion poll. (February 2001). The 33rd annual Phi Delta Kappan/Gallup Poll of the public’s attitudes toward the public schools. Phi Delta Kappan, 82 (6), 468-473.


Moses announces new testing program. (1999, September) TEPSA, 56, 5.


National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs (NCELA) (2005).


Neufeld, S. (2000, October 2). Backlash fermenting against school tests: Groups organize to complain about STAR. San Jose Mercury News.


Education Week.


Schrag, P. (January, 2000). Too good to be true. The American Prospect.


Shuttlesworth, Bill. (2003-2005). Social Studies Specialist Region VI Education Service Center. bshuttlesworth@esc6.net


Social Studies Center. (2000-2004). Education Service Center VI in collaboration with Texas A & M University at College Station & Sam Houston State University at Huntsville, Texas.


Texas Education Agency (1999). *Texas Student Assessment Program: Student Field Test Performance Results*. Austin, Texas: Author.

Texas Education Agency (2001-2002). *Texas Student Assessment Program: Student Field Test Performance Results*. Austin, Texas: Author.


Texas Education Agency (2000-2004). *Texas Student Assessment Program: Student Performance Results*. Austin, Texas: Author.


Utah Legislature Interim Committee on Education. (May 2005). ASCD SmartBrief. [On-line]. Available: ascd@smartbrief.com


Wiggins, Karen. (2000-2004). Social Studies Center, a Division of Texas Education Agency. Richardson, TX.


