The Dissertation Committee for Dongmei Li
certifies that this is the approved version of the following dissertation:

# Holding Public Schools Accountable on Non-academic Measures:

# Examining Districts' Perceptions and Implementation of the

# 'Community and Student Engagement' Policy in Texas

**Committee**:

_____
Jennifer Jellison Holme, Supervisor

_____
Angela Valenzuela

_____
Mark Anthony Gooden

_____
Kevin Michael Foster

_____
Martha N. Ovando

**Holding Public Schools Accountable on Non-academic Measures:**

**Examining Districts' Perceptions and Implementation of the**

**'Community and Student Engagement' Policy in Texas**

by

**Dongmei Li, B.A.; M.A.; M.Ed.**

**Dissertation**

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

**May 2017**

# Acknowledgements

Growing up in very poor family and being in a non-STEM area in college, I had never planned to study abroad as many other Chinese students have done. Even though I was an English major, writing in English was my least favorite task. Dissertating in English or becoming a researcher was far from my imagination. After I completed my first masters' program in China, I never would have dreamed that I would be enjoying or able to spend the next several years of my life pursuing a doctoral degree in the U.S., or that today I would be standing on the precipice of a Ph.D., and embarking on my academic career at Rice University as a postdoctoral research fellow. There is no way I could have made it by myself, and the number of selfless souls who assisted me on this journey is far too long to list. However, it would be very remiss of me not to acknowledge the people who played a critical role in my growth and development throughout this process.

I first want to thank the members of my dissertation committee. I took classes or did independent study with all of you and learned a great deal about the subject content, as well as the content of your characters. I learned from all of you what it means to bring passion and effort to scholarship, teaching, and service, which will continue to guide my career in the future. In particular, I highly appreciate your willingness to serve my committee, challenging and supporting me from the proposal stage until now. It was a great pleasure to go on this journey with all of you.

Next are the EDA professors, UT and off-campus supervisors I worked with. It is

such a blessing that I have been able to find work that I love, and that I work with people that I love and respect as well.  It has been an absolute pleasure for me, an international student, to work with you and learn from you.  And I hope to experience such enjoyable and supportive workplace environments in my future career.  Thank you, Dr. Gooden and Dr. Saenz, for hiring me as your teaching assistant for several semesters.  Dr. Jabbar, thank you for giving me the research assistantship, Stata training, and publishing opportunities.  Dr. Holme and Dr. Jabbar, I am glad I was of assistance to your teacher turnover project.  Dr. Somers, thanks for the teaching assistantship; and Matt, I thank you for the data work training and your care about my dissertation progress and well-being when I was your research assistant.  To the OnRamps people, I am thankful to your support of my work there.  Mary, working at your Capitol office is always an invaluable experience.  I thank you and Dr. Foster for giving me the education policy fellowship opportunity.  Frances and Kristie, I am fortunate to be part of the CPPP family. Thanks for your support of my professional growth there and giving flexibility to my dissertation writing.

To the Department of Educational Administration.  The EDA conference travel funding helped me tremendously.  The departmental funding for Jackson Scholars took me to four conferences in two years without financial burdens.  The EDA staff are professional and supportive to graduate students: Professor Holme and Professor Valenzuela, our current and previous graduate advisors, and Hortensia, Jacinda, Lizbeth, and Ryan.  I am extremely thankful to the tremendous support of Professor Holme, my

dissertation committee chair.  I am so fortunate to have you with me on this journey.
You exemplify what an amazing chair looks like, which becomes my professional goal
should I become a faculty member in the future.

To all external mentors.  Dr. Lindle from the Jackson Scholar Program and Dr.
Mansfield from The William L. Boyd National Educational Politics Workshop.  Dr. T.C.
Chan, we met at an AERA conference a few years ago; thanks so much for supporting
my research and dissertation.  You all have your family and work responsibilities, but
you are always there when I need your mentoring.

To my parents.  Unlike some parents in the countryside who had their daughters
drop out of school, you insisted I went to high school, and from there I started my college
journey and transformed my life in a way I could have never imagined.

Finally, to my partner CJ who has been together with me for more than a decade.
You took me to different countries to see the world and helped me realize my great
interest in educational studies.  You have been by my side through each step of this
journey.  Without your encouragement and support, I would have never come to the U.S.
to embark on this doctorate journey.  You make me a more confident person, and you
helped add a new researcher to the field of education policy.

**Holding Public Schools Accountable on Non-academic Measures:**

**Examining Districts' Perceptions and Implementation of the**

**'Community and Student Engagement' Policy in Texas**

Dongmei Li, Ph.D.

The University of Texas at Austin, 2017


SUPERVISOR: Jennifer Jellison Holme

Abstract: As resistance to high-stakes testing has grown across the country, some states have experimented with non-testing based and local models of accountability reform.  Texas is one such state, which implemented the 'Community and Student Engagement' (CASE) policy.  This policy maintains current testing, but also requires districts to self-evaluate in compliance with statutory reporting requirements.  The eight required areas include fine arts, wellness and physical education, community and parental involvement, the 21st century workforce development program, the second language acquisition program, the digital learning environment, dropout prevention strategies, and educational programs for gifted and talented students.

To date, however, little work has thoroughly examined new reforms like CASE. Meanwhile, the Every Student Succeeds Act (ESSA) grants states flexibility towards assessment and takes full effect in the 2017-18 school year.  Without an understanding of

current reforms, we will lack the empirical knowledge to inform states how to best access the ESSA opportunity to improve their accountability systems. To remedy this gap, this project examined CASE, a new policy added to Texas' accountability system in 2013.

The purpose of this study was to understand public school districts' perceptions and implementation of CASE. Using mixed methods, including collecting survey data from one-third of the districts across the state and conducting semi-structured interviews with a subsample of district leaders, I explored how school district leaders perceive the CASE policy and its impact, and how they implemented the policy. The analysis of survey data shows an overall support of CASE by most district leaders and different patterns of perceptions by district type. The results from survey and interview data indicate varying policy implementation and impact patterns in different districts.

This study will contribute to a better understanding of local districts' perceptions and implementation of non-academic accountability reform. It provides timely empirical evidence for Texas and other states to revise their accountability systems and inform respond to the new ESSA requirements of measuring non-academic domains along with academic measures.

# Table of Contents

# List of Tables

# List of Figures

# CHAPTER 1 INTRODUCTION

**The Prevalence of High-Stakes Testing Based Accountability**

Test-based accountability has become a worldwide trend in education (Loeb & Figlio, 2011). Many countries test students and hold teachers and schools accountable for testing outcomes. In China, a country with a long tradition of heavy testing, high-stakes test scores are not only the metrics for students' advancement from elementary school to secondary school to college, but they also determine a school's reputation. In England, seven and eleven-year-old students have taken standardized tests ever since the 1988 national curriculum was put in place (Alexander, 2011). Schools in England are also required to report accountability results from test scores (Burgess, Propper, Slater, & Wilson, 2005). Likewise, Chile began reporting school aggregate testing scores two decades ago (Mizal, Romaguera, & Urquilola, 2007). In 1988, Chile implemented a new test system that publicly disseminated results; the government also utilized those scores to allocate educational resources and even considered using the scores to advance accountability and transit incentives (Mizal, Romaguera, & Urquilola, 2007). Also, Australia implemented its new accountability system in the late 2000s (Loeb & Figlio, 2011).

In the case of the United States, federally mandated high-stakes testing has gained dominance over the last ten years. The No Child Left Behind Act (commonly known as NCLB), a reauthorization of The Elementary and Secondary Education Act (ESEA), was passed in 2002; thereafter, the federal government mandated that a test-based accountability system be established in all states (Elmore, 2002; Loeb & Figlio, 2011).

1

Prior to 2002, fewer than half the states had accountability initiatives (Elmore, 2002).

NCLB heralded a stringent and inflexible education accountability system. It increased

testing requirements and established demanding standards that targeted all students and

subgroups (Linn, Baker, & Betebenner, 2002). This system mandates when and how

often mathematics, reading, and science must be tested for specific grades and that

Adequate Yearly Progress (AYP) be met and reported. In addition, NCLB contained

specific sanctions for chronically low-performing schools based on student test scores.

Previous studies have explored the impact of test-based accountability and have

reported mixed findings in terms of student outcomes. In China, a common critique of

the excessive focus on testing is that Chinese students are said to have high test scores but

"low ability," with many ill-suited to work in foreign companies (Zhao, 2010).

Numerous studies report that Chinese students are under constant and substantial physical

and psychological pressures and they often sacrifice sleep and recreation to study (Ding

& Hu, 2004; J.L. Li, 2002; J. Li & Xu, 2011; Liu, 2013; Luo & Chen, 2013; Wen, 2007;

Zhou, 2001). Research about England's secondary public schools finds the

accountability measure of partial student achievement incentivize schools to improve the

performance of students on the margins, while reducing low-achieving children's

educational gains and test performance (Burgess, Propper, Slater, & Wilson, 2005).

Other research generates the same mixed results. On the one hand, international data

show improvement in reading, math, and science among UK primary students (Alexander,

2011). On the other hand, this same study finds reading gains deprive students of reading

enjoyment, and the achievement gap between high- and low-performing students does not

change regardless of the presence of test-based accountability (Alexander, 2011). The United States is no exception. NCLB has been implemented for more than a decade and research findings conclude that it has been largely ineffective in reducing the achievement gap and improving the academic performance of the most disadvantaged students (Holme, Richards, Jimerson, & Cohen, 2010; McCarty, 2009; McNeil, Coppola, Radigan, & Vasquez Heilig, 2008; Menken, 2010; Ravitch, 2010; Rumberger & Gándara, 2004; Vasquez Heilig, Young, & Williams, 2012).

**Growing Dissatisfaction with NCLB**

Dissenting voices at the grassroots level have existed since NCLB's inception, and increasingly accounts show that resistance to high-stakes test-based accountability has spread to several states, involving a broad spectrum of participants (Schaeffer, 2002; Strauss, 2012b). The largest teachers' union, the National Education Association (NEA), opposed NCLB accountability provisions of standardized testing, due to its penalizing of teachers and other educators (Williams, 2006). The NEA continued its anti-testing stance and has approved similar resolutions opposing NCLB (Strauss, 2012b). The second-largest teachers' union, the American Federation of Teachers (AFT), adopted a resolution stating that standardized testing has undermined education (Schaeffer, 2002; Strauss, 2012b), and continued its critical attack on high-stakes testing until 2014 (Strauss, 2014b).

Various educational associations and school boards also oppose high-stakes testing. FairTest organized several groups to launch the national resolution on high-stakes testing (Strauss, 2012b). Many parents, students, teachers, principals, school board members, and researchers from across the nation endorsed similar resolutions

3

(Strauss, 2012b). Oregon and Massachusetts NEA affiliates proposed a three-year moratorium on all testing. A parent advocacy group, New York State Allies for Public Education, aimed to encourage the parents of 250,000 students to boycott New York state tests in 2015 (Reid, 2014). The Florida School Board Association has passed its own resolutions (Strauss, 2012b). Representing Florida's 67 school boards, this association called upon the state to only use test results diagnostically rather than for other purposes (Solochek, 2014). In Texas, over 800 school boards passed resolutions that stated tests were "strangling education" (Strauss, 2012a, 2012b). The Colorado Education Association joined other organizations to demand that Colorado withdraw the Partnership for Assessment of Readiness for College and Careers and suspend high-stakes standardized tests for three years (Ravitch, 2014a). Two more Colorado school boards passed resolutions in early October of 2014, requesting the state to grant a five-year waiver from two state tests (Engdadl, 2014).

Even teachers have taken action against annual student assessments (Strauss, 2014b). In spring 2013, Seattle school teachers boycotted testing, which attracted national attention (McCauley, 2013). In 2014, two first-grade teachers from Tulsa, Oklahoma notified their students' parents that they would not administer any standardized tests (Hagopian, 2014). In February of that year, teachers from a Chicago school serving pre-k to eighth grade students voted to refuse to administer the state mandated test (Hagopian, 2014). In May of 2014, New York City teachers from a high school serving English language learners announced that they "would refuse to

administer a test that was culturally and linguistically inappropriate for their students"

(Hagopian, 2014).

Many parents across the country have also protested by opting their children out

of high-stakes standardized tests (Bidwell, 2014b).  A 2014 November news article also

showed many New Mexico families opted their children out of state testing (Nott, 2014).

By February 2015, over 300 Albuquerque students had opted out (Claytor, 2015).  The

Coalition for Better Education called for donations to place two billboards on Colorado's

major highways to inform parents of their opting-out rights (Ravitch, 2014e).  In Texas,

the group "Texas Parents Opt Out" wanted a moratorium on state-mandated testing, with

thousands of Texas parents opting their children out of the State of Texas Assessments of

Academic Readiness (Denny, 2014).  Additionally, compared to the previous year, more

Florida parents opted their children out during the testing month in 2015 (Solochek,

2015).  In New York, the number of parents refusing to allow their children to take state-

required high-stakes tests increased from a few thousand to 60,000 between 2013 and

2014 (Hagopian, 2014; Strauss, 2014b).  By early 2016, twenty percent of the eligible

students in New York State refused to take the state's accountability test (Strauss, 2016).

This percentage is four times the non-testing population permitted by the Every Students

Succeeds Act (ESSA), which may sanction states with less than 95 percent of the eligible

students taking tests by withholding federal funding (Strauss, 2016).

Some students from certain states have resisted in their own ways.  In March 2013,

a Colorado junior organized about 400 students from eight schools to protest state

standardized testing (Steele, 2013).  In April of 2013, two Cleveland High School

students organized a walk out that included over 70-students to protest the state-mandated science test for juniors (Dungca, 2013). In the same month, more than 300 students from over 25 different Chicago public schools walked out to boycott the second day of the state-mandated test (McCauley, 2013). In November 2014, Colorado high school seniors refused to take the state-mandated testing (Reuter, 2014); approximately 200 from Boulder opted out (Garcia, 2014) and walked out of their classrooms (Brundinnov, 2014). Out of 538 seniors at one Colorado school, nine students took the science test and ten took the social studies test (Ravitch, 2014c) and 30 percent of seniors at a different school submitted test refusal (Garcia, 2014). 97 percent of the seniors at another Colorado high school did not show up on their testing day (Ravitch, 2014d). Taken together, over 5,000 Colorado seniors refused to take state-mandated standardized test in 2014 (Hagopian, 2014).

**Changes to the High-stakes Testing System**

The momentum for weakening high-stakes testing peaked in 2014, called "the greatest year of revolt against high-stakes testing in US history" (Hagopian, 2014). High-stakes testing resistance not only occurred in many more states, but also continued to lead to changes in some states at the local district levels (Strauss, 2014b). In August 2014, Florida's Lee County school district voted to abolish all testing and obtained opt-out approval from the District Board of Education (Mathews, 2014; Strauss, 2014a). A week after this initial decision, while the district reversed its opt-out decision (Bidwell, 2014a), the district did drop all district-required exams (Strauss, 2014b). Miami-Dade County also dropped a set of district interim exams (Alvarez, 2014). Elsewhere, the Oklahoma

Education Association united parent groups to "soften" third grade promotion tests (Strauss, 2014b).  The Pittsburgh School Board reduced K-5 testing by half (Ravitch, 2014b).  In New York, about 20 school districts did not administer tests with experimental items for the 2015 state testing period (Strauss, 2014b).

As the nationwide revolt against high-stakes testing grows, some state legislatures are seeking to reduce standardized tests.  According to a FairTest September 2014 report, public pressure on high-stakes testing led to reductions in state testing mandates, decreases in score-based consequences, and suspensions of Common Core tests (Guisbond, 2014).  States such as Alaska Minnesota, Rhode Island, and South Carolina cancelled or postponed graduation testing, while other states and cities–North Carolina, Oklahoma, Texas, Virginia, and New York City–decreased the number of required tests (Guisbond, 2014).  For instance, anger over an overwhelming number of state-mandated tests pushed parents to demand changes during the 2013 legislative session in Texas (Serrano, 2013).  The legislature finally passed House Bill 5 (HB5), which is a comprehensive bill covering academic and financial accountability.  According to Representative Jimmie Don Aycock, this bill was a response to nearly two years of pressure emanating from educators and parents (Serrano, 2013).  HB5 reduced the number of end-of-course (EOC) exams required for high school graduation from fifteen to five and ended the proposed requirement that EOC exam scores count as fifteen percent of a student's grade.  The new bill also offers multiple graduation plans and aims to give districts local control over graduation requirements (House Research Organization, 2013).

Some states' legislatures started tackling the removal of certain high-stakes

testing in 2015. The Providence Student Union from Rhode Island, called "one of the

most organized and creative student groups in the nation in opposition to high-stakes

testing," forced the state legislature to vote for a three-year moratorium on using high-

stakes (Hagopian, 2014). Connecticut worked to de-emphasize the role of testing in

rating schools for the accountability system. The new assessment system intends to go

beyond test scores to include "the arts, civics, physical fitness, attendance, and even

qualities such as student persistence and personal development" (Megan, 2014). In mid-

January 2015, Pennsylvania State Senator Andy Dinniman stated he would introduce

legislation to terminate the use of the high-stakes Keystone Exams for a high school

graduation requirement (Ravitch, 2015). The Mississippi House unanimously passed HB

665, which requires four subjects to be tested as part of the graduate exam; passing is not

required for graduation (Amy, 2015). In the same month, the Georgia House passed HB

91, which removes the graduation test requirement (Redmon, 2015).

**Experimenting with Alternative Models of Accountability**

Partly due to increasing pressure from the public, some states have pushed back

against federally-defined measures and modified their accountability systems. California

and Texas, two of the nation's most densely populated states, adjusted their

accountability systems to incorporate non-academic measures that are important to local

stakeholders and experimented with local models of education accountability. These

states have previously set the precedent for major changes in national education. Their

practices merit attention, as they might serve as a model for other states seeking

alternative assessment mechanisms and spearhead new trends for educational accountability.

California modified its educational accountability systems by linking its new finance formula to education accountability plans. In 2013, the state passed the Local Control Financial Formula (LCFF) that includes an important component: the Local Control and Accountability Plan (LCAP). The LCFF was implemented in January 2014, under which local education agencies are to develop a related budget and a plan to achieve eight state priorities (Sacramento County Office of Education, n.d.). This legislation changed how California allocates funding to school districts, and how they support and intervene in underperforming districts (California Legislative Analyst's Office, 2013). Now, using the state-approved LCAP template, local educational agencies (LEAs) adopt their budget plans and develop annual goals and actions to meet eight state priorities (Kerchner, 2014). According to the LCAP statute, each district must use both quantitative and qualitative measures and seek community input to develop the accountability plan (Vasquez Heilig et al., 2014). After the LCAP is adopted by the district, it moves to the county education office for review and approval, after which it goes to the State Superintendent for final approval (Vasquez Heilig et al., 2014).

Under California's LCFF and LCAPs, test scores cannot be the single indicator used in the accountability system (Kerchner, 2014). This new evaluation mechanism requires school districts to report progress in seven areas in addition to student achievement: student outcomes from other exams, implementation of the Common Core, school climate, student engagement, parental engagement, course access, and basic

9

services (Kerchner, 2014).  This system is considered more flexible because it allows school districts to add their own goals and priorities (Kerchner, 2014).  Districts can measure self-identified specific goals in their own ways (Vasquez Heilig et al., 2014). The seven California districts that received NCLB waivers created their own multiple-measure accountability index; in other places, the LCAP followed the state template (Kerchner, 2014).

Texas is the other large state to experiment with alternative measures to its test-based accountability system.  In 2013, the 83rd Texas legislature passed HB5, which added two non-academic provisions–financial accountability and community and student engagement (CASE)–into the state's existing test-based accountability system. Proponents argued that HB5 would broaden educational accountability criteria and decrease test score reliance (House Research Organization, 2013).  Supporters also noted that new rating categories of CASE and financial accountability would demonstrate a better overall understanding of school and district performance (House Research Organization, 2013).  Implemented immediately after its passage, the evaluation of districts' performance on CASE unfolded during the academic year of 2013-2014.  It is unknown how school districts have perceived and implemented the CASE policy and this dissertation takes up this issue.

**Community and Student Engagement Under HB5**

This dissertation examines the local perceptions and self-evaluation of CASE as embodied in HB5.  This new policy establishes a rating system of eight areas and one compliance factor: (1) fine arts, (2) wellness and physical education, (3) community and

parental involvement, (4) twenty-first-century workforce development programs, (5) second language acquisition programs, (6) digital learning environment, (7) dropout prevention strategies, (8) educational programs for gifted and talented students, and (9) compliance with statutory reporting requirements. The statute does not allow the Texas Education Agency (TEA) to determine the rating criteria of those nine factors. Rather, it requires public school districts to set up a local committee define their own criteria and creating their own scoring rubric for measuring the required areas. Each school district must assign ratings of exemplary, recognized, acceptable, or unacceptable for both districts and campuses for the first eight factors. A rating label of "yes" or "no" is assigned to factor nine. Districts report their established first-year evaluation ratings to the TEA no later than August 8 of each year starting 2014. In so doing, they are supposed to showcase areas of excellence and identify areas of improvement in engaging students and the community.

The development and passage of CASE, which was incorporated in HB5, was a top-down political process. The aforementioned Representative Aycock, the Public Education Committee Chair and leading author of HB5, was critical to the proposal and passage of CASE. HB5 is his policy priority and he is a big proponent of local control. Aycock was reported to state his penchant for local flexibility: "I'm somewhat willing to let local districts say, 'This is where we need to get to as a community,' and let them say, 'How did we do?'" (Michels, 2013). He also says that his education-related experiences–serving on school and college boards–made him think about improving public education (J. Aycock, personal communication, October 17, 2014). He claimed that the intention

behind CASE was to foster and stimulate conversation between educators and the community (J. Aycock, personal communication, October 17, 2014). Additionally, Aycock argued that CASE factors were of great significance for student engagement and will keep students interested in schooling (J. Aycock, personal communication, October 17, 2014).

Another important legislative actor is Representative Joe Deshotel, the co-author of HB5. Deshotel originally filed a separate bill, HB1423, in response to his constituents' needs. The Jobs for Texas Coalition, the Texas Chemical Council, and several other professional organizations had approached his office with alternatives to high-stakes testing (M. Quevedo, Representative Deshotel's legislative director, personal communication, October 17, 2014). Due to duplicate filings by two offices, HB1423 was later incorporated into Aycock's bill, which was ultimately developed into HB5. Prior to the opening of the 2013 legislative session, Aycock sought inputs from superintendents and other educators. He brought the idea to the Public Education Committee during the session and the Committee staff collectively drafted the nine measures (J. Aycock, personal communication, October 17, 2014).

The CASE policy is significant because it requires districts to organize a local committee that encourages them to involve parents, business representatives, and the local community. It gives each school opportunities to demonstrate the strength of their programs that are not reflected on standardized test scores, recognize areas in need of improvement, and set future goals that are valued in the community. As proponents argued, this new policy would allow local districts to go beyond test scores and present a

more comprehensive picture of campus and district performance (House Research Organization, 2013). In addition, supporters asserted that the bill would strengthen public investment in the public education system by involving local groups of parents, business leaders, and other community members in deciding the best practices for school evaluation (House Research Organization, 2013).

However, the CASE policy has some noteworthy issues. It does not replace the current high-stakes testing-based accountability system and merely adds new, locally-determined measures. The self-evaluating nature of the policy makes it lower stakes and might be perceived as less valid in the eyes of the pubic. Its locally developed rating criteria in the required areas vary across districts, which makes ratings incomparable across the state. The extent of local stakeholders' cooperation might be a problem given the varying capacities of each school district and community (Lopez & Valenzuela, 2013). It is also unclear to what extent the unfunded policy will enable low income districts to make improvements in their community engagement.

**Purpose statement and research questions.** The purpose of this exploratory study was to understand public school districts' perceptions and implementation of the CASE policy. To fulfill this purpose, I answer the following research question: how do public school district leaders perceive the CASE policy and its impact, and how have they implemented the CASE policy? Three sub-research questions necessarily extend from this primary concern, including:

1. How do public school district leaders perceive the CASE policy?
2. How have public school districts implemented the CASE policy?
   a. Who was involved in determining the CASE rating criteria?

b. What was the process for determining the CASE rating criteria?

3. How do public school district leaders perceive the impact of the CASE policy?

**Significance of this study**.  In an era of increasing resistance to high-stakes testing, it is vital to understand new accountability reform.  Texas is one of the implemented examples of a state which moved educational accountability beyond test scores to experiment with locally defined non-testing measures.  Many states are dissatisfied with the narrow measures mandated by the NCLB system, and potentially may move toward accountability reform.  To date, however, little work has thoroughly examined these assessment reform emerging, in particular, in Texas and California as a response to the federal mandate and local dissatisfaction with it.  Meanwhile, the ESSA grants states flexibility towards assessment and takes full effect in the 2017-18 school year.  Without an understanding of current reforms, we will lack the empirical knowledge to inform states how revise their assessment system and how to best access the ESSA opportunity to improve their accountability systems.

To remedy this gap, this study examines local districts' perceptions and implementation of Texas' accountability policy, the 2013 CASE policy.  This study generates empirical evidence on local perceptions and implementation of non-academic accountability reform and will lead to a better understanding of accountability reform that empowers local districts.  It informs possible policy modifications and adds new knowledge to the existing literature on accountability.  It will also contribute to the policy debates about the revision of accountability systems in many states and the national trend of assessing the non-academic indicators as required by the ESSA.  The ESSA requires

schools to assess at least four indicators, that is three academic indicators and one school quality indicator (Klein, 2016). Each state may opt for "student engagement, educator engagement, access to and completion of advanced coursework, postsecondary readiness, school climate and safety, or whatever else the state thinks makes sense" to assess their school quality (Klein, 2016; Korte, 2016). The factors of CASE (p.12) match some of the descriptions of school quality. Thus, the findings of this study will provide timely information for Texas as well as other states to effectively define and measure the required non-academic indicator.

# CHAPTER 2 LITERATURE REVIEW

In order to set the stage for my study of Texas's implementation of the local accountability model, I first provide a brief overview of education accountability, from the early twentieth century to the Every Student Succeeds Act (ESSA). Second, I trace the history of Texas accountability from its instantiation to the present accountability model, characterized by academic measures, community and student engagement, and financial management. Third, I review the impacts of high-stakes testing before and after NCLB on students, schools, and teachers, and then discuss alternatives to test-based accountability found in both literature and practice. Finally, I describe the conceptual frameworks that guide this study.

**History of Education Accountability in the U.S.**

Understanding the current push to broaden assessment measures and increase local input into educational accountability in Texas and other states requires a review of how schools have historically been held accountable. In the U.S., public education has traditionally been the terrain of local and state governments (Jefferson-Jenkins & Hill, 2011). However, since the passage of ESEA in 1965, the federal role in public schools has increased. This section aims to present how public education has shifted from local control to state and federal control. I first offer a review of the period preceding 1940 and then discuss the era of 1940‑80. After focusing on the following two decades, 1980‑2000, I explore the history of NCLB, which is the dramatic federal control of national public education. Finally, I introduce and briefly analyze the ESSA, the reauthorization of NCLB and the present form of ESEA of 1965.

**Before 1940**.  Local control used to be a key characteristic of American public

school governance (Howe & Meens, 2012).  In most states, the local community first

established schools and controlled them directly (Dennis, 2000).  From colonial times to

the early nineteenth century, children received schooling from institutions and

arrangements that were largely supported by local entities (Dennis, 2000).  For example,

the 1785 and 1787 Land Ordinance and Northwest Ordinance required a locally-funded

system of township schools (Jefferson-Jenkins & Hill, 2011).  As Labaree writes:

> During most of the nineteenth century, the local school was the primary unit of
> educational governance for most Americans.  An individual community built a
> school, hire a teacher, raised money through local taxes and fees, and
> implemented education on its own terms.  Outside help was neither offered or
> welcome…even in large cities, control of education tended to rest at the ward
> level (Labaree, 2000, p.29).

Indeed, education was under local control and decentralized at that time as the rural

school was an integral part of the community, though less structured and not compulsory,

and even laymen were part of the educational decision making process (Tyack, 1974).

Additionally, early urban education governance followed a rural model (Tyack, 1974)

that was "haphazard and relatively informal … [and] fostered a close relationship

between school and community… not only in a legal sense but in a social one as well"

(Kantor, 2001, p.319).

In the late nineteenth century, urbanization and industrial development intensified

and foreign immigrants flooded into the U.S. (Tyack, 1974).  Villages "grew into

congested, heterogeneous cities [and]…decentralized decision-making…struck education

leaders as anarchy" (Tyack, 1974, p.39).  As a result, education reformers of the era

sought a uniform, centralized, and standardized system.  Education centralization and

standardization became important.  Another key development in the centralization trend

was the "consolidation of small local districts into larger, townwide systems" (Kaestle &

Smith, 1982, p.384).  Urban education governance gradually replaced village schools

with a corporate model of centralized governance and specialized administrative

structures, followed by an increased growth of layered bureaucracies (Tyack, 1974).

Education structural reformers intended to create an insular system in which

responsibilities were confined to schools and professionals who were given due

autonomy based on their capability (Tyack, 1974).  Professional educators replaced

laymen and took a critical role in the educational process (Tyack, 1974).  Students and all

other members within the educational system were supposed to be accountable for

particular prescribed duties (Philbrick, 1885, as cited in Tyack, 1974).

In the meantime, state governments gradually became more involved in regulating

local education (Kaestle & Smith, 1982); this was met with some objection by localities.

Local complaints and resistance arose when Massachusetts established a state board of

education in 1840 and Wisconsin attempted to consolidate school districts in 1865

(Kaestle & Smith, 1982).  Local control advocates in Connecticut and Ohio also resisted

their states' consolidation efforts (Kaestle & Smith, 1982).  Nonetheless, by 1860 most

people from the northern U.S. accepted this centralization trend, and, by the early 1900s,

a majority of southerners endorsed the idea as well, owing to the "rhetoric of progress

and equality" that would be brought by states' involvement in regulating public schools

(Kaestle & Smith, 1982, p.387).  Incrementally, state governments became primary actors

in local K-12 education through increasing curriculum regulation, teacher certification control, and governments' shares in local education budgets (Kaestle & Smith, 1982). This change was historical, as it marked a time when state control and regulation stepped into the public education arena, which used to be exclusively under local control.

Federal involvement in public education began before 1940 (Kaestle & Smith, 1982), but played a minimal role during this period. For example, in the nineteenth century, the federal government provided vocational training and land grants (Jefferson-Jenkins & Hill, 2011; Kaestle & Smith, 1982). The federal Office of Education was established in 1867 to collect schools' and teachers' information, enabling states to develop effective schools (Jefferson-Jenkins & Hill, 2011). Fifty years later, the 1917 Smith-Hughes Act offered federal aid to promote high school vocational education (Jefferson-Jenkins & Hill, 2011; Kaestle & Smith, 1982).

**1940 to 1980.** The Second World War accounted for much of the significant expansion of federal intervention in public education (Jefferson-Jenkins & Hill, 2011). The 1941 Lanham Act and the 1950 Impact Aid laws provided school districts with payments that alleviated communities' burdens from military presence and other federal installations (Jefferson-Jenkins & Hill, 2011). The 1946 George-Barden Act provided high schools with funding for vocational training in agriculture, industry, and home economics (Jefferson-Jenkins & Hill, 2011). Crises after World War II strengthened federal intervention in public education. In response to the Soviet launch of Sputnik, the U.S. Congress passed the National Defense Education Act (NDEA) in 1958 (Jefferson-Jenkins & Hill, 2011). Yet, the federal role in developing education policies was still

minor prior to 1960, remaining marginal throughout the end of the 1950s (Kantor, 1991).

Thus far, federal support in vocational and teacher education training "did not threaten

local interests and required minimal federal regulation" (Kantor, 1991, p.48).

Federal involvement in public education dramatically expanded in the 1960s

(Jefferson-Jenkins & Hill, 2011; Kantor, 1991).  President Lyndon Baines Johnson's War

on Poverty led to the Elementary and Secondary Education Act (ESEA) of 1965, which

increased the role of the federal government in educational decision-making.  The

comprehensive ESEA contains six sections, including federal aid to low-income students

(called Title I), school instructional materials, educational services, and educational

research and training.  ESEA funds national primary and secondary education but forbids

establishing a national curriculum (Jefferson-Jenkins & Hill, 2011).  This was the federal

government's "answer to the call for greater attention to the quality of America's schools

and the needs of students from less advantaged homes" (Nichols & Berliner, 2007).  To

comply with ESEA, local education agencies only needed to report to the state

government and the U.S. Office of Education regarding the development of

compensatory educational programs (Richburg, 1971).

ESEA specifically supplied federal aid to districts and schools serving large

numbers of poor children (The Network for Public Education, 2015).  This resource

allocation was through the Title I program and aimed to bridge the gap in core subjects

between low-income and middle-class students (Farkas, Hall, Finn, Carnine, & Meeder,

2000).  This act also implemented the Title I Evaluation and Reporting System, which

suggested that students be tested in the fall and spring of each academic year and only

required the reporting of test data (Linn, 2000). Such accountability testing was largely used to evaluate program effectiveness. Testing was not federally mandated, but a function belonging to states (The Network for Public Education, 2015). Over the years, other programs such as Title II and aid for homeless students were added, which also supported low-income students.

The original intentions of this law were to increase educational inputs to schools and districts that enrolled high proportions of poor students; it did not contain any reference to testing (The Network for Public Education, 2015). Basically, public schools were more accountable for regulation compliance as a result of ESEA (federal financial) inputs. According to Dyer (1971), the 1965 ESEA did not include corrective or punitive guidelines and was more of an accounting procedure rather than an accountability measure. In a similar vein, Megan (2014) noted that expenditure monitoring and compliance were the bases of public school performance when the accountability mechanism came into being before the 1980s.

Between 1965 and 1980, federal policy makers focused on basic skills, resource distribution, and increased parental involvement in Title I schools (National School Boards Association, n.d.). Accountability was low stakes and there were no requirements of severe sanctions despite more conditions being attached to schools receiving federal aid. The federal government was involved in a small way, holding more public schools accountable for educational inputs rather than outputs. In the 1970s, some states required a minimum competency exam for grade promotion or high school graduation (Champion, 2007). From 1973 to 1983, the number of states involved in standardized testing

increased from two to thirty-four (Linn, 2000), but tests were focused on basic skills (Johnson, 2004; Linn, 2000).

**Between 1980 and 2000.** While the 1980s and 1990s witnessed a growing influence of the federal government on the national education landscape, accountability still remained largely at the state level. The federal government established the Department of Education in 1980 and entrusted the Educational Testing Service to administer the National Assessment of Educational Progress in 1983 (Jefferson-Jenkins & Hill, 2011). Garland (2014) wrote in a news report that, before the publication of *A Nation At Risk* (ANAR), the federal government played a minimal role in public education; President Ronald Regan even wanted to have the Education Department abolished. When the 1983 ANAR by the National Commission on Excellence in Education warned that the U.S. no longer held the educational edge when compared to other countries (Jefferson-Jenkins & Hill, 2011), there was a concern that American schools were not as good as those in other nations (Nichols & Berliner, 2007). According to Nichols and Berliner (2007), the publication of ANAR not only made "American education the scapegoat of certain social problems," but also "sparked a renewed interest in and urgency about how we educate America's students. That sense of urgency set in motion a series of policy initiatives aimed to improve the American education system," among which was "the call for more consequential testing" (Nichols & Berliner, 2007, p.4).

Ever since the ANAR report, the federal government has become increasingly involved in public education and states have tightened their educational standards,

increasing their control over public schools. The federal government became one of the main drivers of reform (Garland, 2014). ANAR called for tests to go beyond basic skills and measure more rigorous curricular standards (Johnson, 2004). It led to a shift of testing from the minimum competency testing to education excellence. In the midst of the national climate of emphasizing advanced skills rather than basic skills, some states developed statewide standardized tests that were more rigorous than those of the past. For instance, in response to ANAR, between 1980 and 1984, Texas started the Texas Assessment of Basic Skills, which changed to the Texas Educational Assessment of Minimum Skills in 1985 (Haney, 2000). Connecticut also responded to ANAR and had the Connecticut Mastery Test in place for decades to "chart the progress of students and the success of schools" (Megan, 2014).

The period from the 1990s to early 2000s saw standards-based reforms and large-scale accountability assessments, wherein schools were increasingly being held accountable by states for educational outcomes. Educational accountability began shifting its focus from educational inputs to outputs when then Arkansas Governor William ("Bill") Jefferson Clinton introduced the idea of performance-based accountability in the mid-1980s (Elmore, 2002). In 1989, President George H.W. Bush organized an education summit with a group of governors (Loeb & Figlio, 2011; Ravitch, 2010). During this meeting, federal and state leaders reached a consensus of six national goals to be accomplished by 2000 (Ravitch, 2010). Since then, public schools have no longer been held accountable solely for educational inputs, but also for outputs. Between 1989 and 1992, President Bush introduced a series of education initiatives, including the

America 2000 education reform program and an attempt to establish national standards and a national assessment system (Jefferson-Jenkins & Hill, 2011). Governors, corporate executives, the first Bush administration, and the subsequent Clinton administration all wanted measurable results (Ravitch, 2010). This accountability operated in the form of the "horse trade" in which local districts and schools had to produce more academic accountability in order to receive (from states) flexibility in instruction in terms of decision-making (Elmore, 2002). In the 1990s, accountability was "the watchword of public officials and business leaders" (Ravitch, 2010, p.149).

During President Clinton's administration, the performance-based accountability movement gained a tremendous boost. Clinton signed The Goals 2000: Education America Act into law on March 31, 1994. This act called for national tests in language arts, math and science, and national standards (U.S. Department of Education, 1991, as cited in Champion, 2007). It established a list of national goals to be met by 2000 and created voluntary national achievement standards for various academic subjects. States could apply for the grants appropriated by Congress to improve schools. In response, all states except Iowa and Nebraska developed curricular standards and implemented test programs (Johnson, 2004). In October of the same year, President Clinton reauthorized and renamed ESEA as the Improving America's Schools Act (IASA) (Elmore, 2002; Thurlow, 2009). IASA's goals were consistent with the framework of Goals 2000. It amended the Title I act and enforced requirements to schools receiving Title I grants. Title I amendments required that states create performance-based accountability systems for Title I schools (Elmore, 2002). The federal government added mathematics and

English language arts standards for Title I schools to assess students' academic progress and demonstrate accountability (National School Boards Association, n.d.). Further, Title I schools were supposed to achieve the progress goals of that accountability system by 2000 (Elmore, 2002). However, accountability results would not be a condition of receiving federal money, which contrasted with 2001 NCLB requirements.

**The NCLB era**. With the reauthorization of ESEA in 2001, also known as NCLB, federal accountability and control significantly increased (Thurlow, 2009). NCLB asks for more education outcomes from public schools and attached consequences to schools and districts if they failed to meet the AYP plan, which was established by each state and approved by the federal government. Prior to the implementation of NCLB, test-based accountability systems operated in fewer than half the states (Elmore, 2002). This was because previous federal policy on accountability was not stringently enforced. NCLB ushered in an era where all states were mandated to build accountability systems on high-stakes assessments (Elmore, 2002; Loeb & Figlio, 2011). It is noteworthy that NCLB does not change the roles of local communities and states in the accountability system. Localities still have less of a say in what they hold schools accountable for (Howe & Meens, 2012). States' control remains strong in that states are required to interpret the design of their own accountability systems, decide what accounts for AYP, and are allowed to take actions for chronically low-performing schools.

NCLB accelerates the policy shift from school inputs to performance outcomes (Hanushek & Raymond, 2005). In the past, policies at the federal and state levels tended to concentrate on resource provision with a focus on disadvantaged students (Hanushek

& Raymond, 2005).  In contrast, NCLB's central component is the performance measure, solely represented by testing (Elmore, 2002).   Within this testing-driven accountability, students' test scores are the only metric for annually evaluating students, teachers, administrators, and schools (Nichols & Berliner, 2007).  NCLB demands that states determine their respective academic standards and testing systems to measure students' educational progress.  Specifically, it requires that states annually test students in grades three through eight on reading and mathematics and test students at least once during grades ten through twelve (U.S. Department of Education, 2004a).  It also requires science testing at least once for grades three through five, six through nine, and ten through twelve (U.S. Department of Education, 2004a).  Further, NCLB had an ultimate goal that all students should achieve proficiency in reading and mathematics by 2014.

NCLB strengthens how schools (including those receiving Title I funding) are held accountable for outcomes by specifying improvement plans and sanctions for persistently low-achieving schools that fail to reach state standards within a certain period. It mandates schools to annually administer standardized tests and publish report cards with student achievement data and demographics (U.S. Department of Education, 2004b). Schools face consequences when they fail to make annual gains and their students receive low scores (Nichols & Berliner, 2007).  The schools attended by low-performing students must provide free tutoring, an after-school program, or send those students to higher-performing schools if their students fail to fulfill their states' accountability standards for more than two years, (Nichols & Berliner, 2007).  When schools fall below state standards for more than two consecutive years, they may be overhauled following an

intervention model prescribed by the federal government (Nichols & Berliner, 2007).

These intervention models are tied to School Improvement Grants (SIGs), created in

2009 under President Barack H. Obama's administration. SIGs are allocated to state

education agencies based on NCLB requirements and include four intervention models:

turnaround, restart, school closure, and transformation (Nichols & Berliner, 2007; U.S.

Department of Education, 2010). Under increased accountability, Title I school teachers

are required to be qualified and if a state does not have the approved assessment system,

Title I schools must take corrective action (U.S. Department of Education, 2004b). If

Title I schools fail to meet AYP for three years, they are required to make restructuring

plans (U.S. Department of Education, 2004b). Basically, NCLB compliance is

compulsory unless a state forgoes the federal funding that accounts for about ten percent

of a school district's budget. Despite this federal mandate, states still have some

discretion in determining the type of assessment and the student test pool (Champion,

2007).

      **The ESSA.** The ESSA is a reauthorization of the 1965 Elementary and

Secondary Education Act. After passing the House and Senate with strong bipartisan

support (Turner, 2015), it was signed into law by President Obama on December 10,

2015 (U.S. Department of Education, 2015). It goes into full effect in the 2017-18 school

year. The scheduled revision of the 2002 NCLB act did not occur in 2007 and the

NCLB's "prescriptive requirements became increasingly unworkable for schools and

educators" (U.S. Department of Education, 2015). In 2010, the Obama administration

"joined a call from educators and families to create a better law that focused on the clear

goal of fully preparing all students for success in college and careers" (U.S. Department of Education, 2015). The ESSA is the Congress' response to the call (U.S. Department of Education, 2015).

The ESSA maintains the old testing schedule and the annual statewide assessments and the competitive programs to evaluate and reward effective educators in high-need schools remain unchanged (U.S. Department of Education, 2015). But states can opt to administer "multiple statewide interim assessments during the course of the academic year that result in a single summative score that provides valid, reliable, and transparent information on student achievement or growth" (ESEA of 1965). States can allow districts to use SAT or ACT as high school achievement test (Dufour, 2016; Klein, 2016). While the ESSA does not change the federal requirement of 95 percent participation in tests, states can develop their own testing opt-out laws and decide what to do with in schools that miss test participation targets (Klein, 2016). According to Fair Test, a national center for fair and open testing, nine states had varying testing opt-out policies by March of 2017 and as of now, several state legislatures are considering bills to allow students to opt out.

This bipartisan bill make several changes. While the Title I funding formula is the same as that in the NCLB, the Title II formula on funding teacher quality contains some changes favoring rural states (Klein, 2016). The new law adds the college- and career-ready standards and the innovative local assessment pilot (U.S. Department of Education, 2015). It allows states to design multiple measures on student performance and to develop their own interventions towards struggling schools (U.S. Department of

Education, 2015).  The ESSA also includes the pre-k program (U.S. Department of

Education, 2015), which is "housed at the Department of Health and Human Services,

jointly administered by the Education Department" (Klein, 2016).  Adopting Common

Core standards is not required but optional (Klein, 2016; Korte, 2016).  The ESSA

removes the NCLB provision of teacher and leader evaluation through student outcomes

(U.S. Department of Education, 2015).

Some researchers argue that the ESSA is essentially an extension of NCLB in that

it reflects the same philosophy and direction and the standardized test-driven system is

punitive in nature (Mathis & Trujillo, 2016).  Regardless, it "rolls back much of the

federal government's big footprint in education policy" (Klein, 2016) and "gives state

governments sweeping authority to design, among other things, teacher evaluations and

school accountability systems," (Burnette, 2016).  It also gives "states more leeway in

how they identify their lowest-achieving campuses and the remedies they prescribe to

help those schools improve" (The Hechinger Report, 2016). However, individual states

still must submit their accountability plans to the U.S. Department of Education for

approval, though the latter has limited oversight authority (Klein, 2016).  States and local

districts are less likely to fundamentally change the existing accountability system due to

their previous heavy investing (Mathis, & Trujillo, 2016).

Under the multiple measurement of ESSA, schools must adopt at least four

indicators into their accountability systems (Klein, 2016).  The three academic indicators

can be student proficiency and English language proficiency for all school levels, student

growth on state tests for elementary and middle schools, and graduation rates for high

schools (Dufour, 2016; Klein, 2016).  The fourth indicator is school quality that is

evaluated at each state's discretion.  A state may use indicators of "student engagement,

educator engagement, access to and completion of advanced coursework, postsecondary

readiness, school climate and safety" (Klein, 2016; Korte, 2016).  Researchers see

multiple measures as "a positive development in that a more comprehensive set of

measures will more validly capture the broader set of cognitive and affective learning

goals of schooling" (Morton & Dalton, 2007, as cited in Mathis & Trujillo, 2016, p.674).

These experts also caution that multiple measurement "is an elastic term that includes an

eclectic variety of elements" and in each state, it "can mean many different things and

thus result in many different approaches" (Mathis & Trujillo, 2016, p.674).  While test

scores and graduation rates weigh more than subjective measures (Korte, 2016),

assigning weights to multiple indicators that construct a composite score is challenging

(Mathis & Trujillo, 2016).

      **Summary**.  The review of the history of the education accountability movement

shows a change of trends with regard to the roles of local, state, and federal governance

in public education.  In general, school accountability has shifted from less local input by

school boards and localities to strong federal control in accountability assessment policies.

While localities and states do have some say on policy, the federal government affects the

overarching direction of accountably policies, drives what is measured, and places a great

deal of pressure on certain tests and subjects that are stated in NCLB requirements and

attached to federal funds.

Widespread frustration with the testing system led President Obama to offer ten states flexibility from NCLB mandates in 2012 if those states promised to improve education in their own ways (Feller & Hefling, 2012). The federal government tried to soften their rigid accountability system through waivers, largely due to the concern of the law's requiring an attainment of 100% proficiency in reading and mathematics by 2014. In 2012, a total of 28 other states and the District of Columbia and Puerto Rico indicated that they would apply for NCLB waivers (Feller & Hefling, 2012). Increasingly since 2012, more states have applied for NCLB waivers. By summer 2014, forty-two states and the District of Columbia sought waivers (Camera, 2014; A. Klein, 2014b). By July 31, 2014, the U.S. Department of Education (DoE) had granted waiver extensions to thirteen states (Camera, 2014). The Department demands that waiver extension-seeking states provide improvement plans for low-performing schools and will allow about ten states to go through "an extra-special, expedited renewals process" for a four-year waiver extension (A. Klein, 2014c).

ESEA waivers seem to move the top-down federal intervention more toward decision-making at state levels. ESEA waiver states develop their own reform plans to "advance all students" achievement by maintaining a high bar for student success, closing achievement gaps, improving the quality of instruction, and increasing equity by better targeting support and resources to schools based on need" (U.S. Department of Education, n.d.). Further, the flexibility from ESEA waivers would grant states additional options to turn around their chronically low-performing schools (U.S. Department of Education, n.d.). However, the federal government is still in total control as it determines which

applicant states the waiver goes to depending on whether the state meets the federal requirements to link teacher evaluations with student outcomes.  Therefore, ESEA waivers can be interpreted as a rhetorical gesture since these waivers do not really put control back into states' hands.

This battle between the federal government and some states sheds light on the strings attached by the federal to NCLB waivers.  For example, the state of Washington "once had a waiver, but lost [it] in April (of 2014) after state legislators refused to enact one of the policies that the U.S. Department of Education wanted–a requirement that teachers be evaluated, in part, by their students' performance on state tests" (Shaw, 2014). This incident not only told the public that Washington was the first and the only state to have NCLB flexibility rescinded by the federal government, but this action also sent a warning message to states on the verge of losing their flexibility (A. Klein, 2014a). Contrary to the experience of Washington, Oklahoma was granted a waiver that had been previously denied, after this state gave up the Common Core state standards and returned to its old math and reading standards (Layton, 2014).  Texas's experience with the ESEA waiver is similar to what happened with Washington.  In January 2015, the federal government rejected a new teacher and principal evaluation system that Texas developed to meet the ESEA waiver requirement (J. Chang, 2015).  However, the Texas Education Commissioner said in February that he would not enforce an educator evaluation system that followed the federal requirement (J. Chang, 2015).  Due to these factors, Texas's NCLB waiver was in jeopardy and may be lost (J. Change, 2015; Hope, 2015).  It seems that state flexibility as a result of ESEA waivers must be approved first by the federal

government. The strong top-down federal intervention in states' education accountability system remains unchanged. Even though the ESSA grants states flexibility towards assessment and accountability, the federal control in public education is still influential. As previously discussed, states have to follow the test schedule and design their opt-out laws allowing five percent of their students to legally opt out high-stakes testing.

**Test-Based Accountability in Texas**

Having reviewed the history of education accountability, the trend of stronger federal control, and the recent minor shift toward supporting state decisions in the form of ESEA waivers, I will now explore the history of Texas's test-based accountability to understand the Lone Star state's testing and accountability history, its relation to national trends, and the evolution to the new local evaluation of non-testing measures since 2013 that in some ways parallels the ESEA waiver.

Texas historically has had a strong state role in holding schools accountable to educational outcomes, particularly test scores. Accountability in Texas occurred approximately at the same time when the federal involvement in public education tightened in the 1980s. In 2013, however, the Texas legislature incorporated local assessment measures in non-testing areas into its accountability system. In order to understand Texas's recent policy shift in expanding this system, it is significant to first understand where the Texas accountability system originated. Among the first of the states to implement statewide accountability (Vasquez Heilig et al., 2014), Texas preceded the national trend of performance-based accountability in the 1990s by about one decade. In this section, I review the (high-stakes) testing history in Texas from the

33

1980s' Texas Assessment of Basic Skills (TABS) to the present version of Texas Assessment of Academic Readiness (STAAR).

**TABS.** Texas began to hold schools accountable with TABS between 1980 and 1984 as the result of a legislative requirement. TABS was the first statewide test to be administered, and it evaluated basic skills in reading, writing, and math for third, fifth, and ninth graders in every school district (Champion, 2007; Ott, 2010). This criterion-referenced test was based on Senate Bill 350 of the 1979 Texas Legislature (Ott, 2010). This assessment was low stakes: if ninth graders failed, they could still graduate but they had to retake the test each year until they passed or graduated (Champion, 2007; Ott, 2010). In 1983, the state began to publish schools' and districts' aggregated test results, which became public information (Ott, 2010). The primary purpose was to collect school-level information (Cruse & Twing, 2000), which was diagnostically used to help students in deficient areas (Ott, 2010). Schools faced no consequences during this time, but TABS ushered in the era "of high-stakes accountability for large-scale assessment in Texas" (Cruse & Twing, 2000, p.328).

**TEAMS**. The 1984 Texas Legislature created the second assessment, which was known as the Texas Educational Assessment of Minimum Skills (TEAMS), implemented in 1985 (Champion, 2007; Haney, 2000; Ott, 2010). The change from TABS to TEAMS was a shift from basic skills to minimum skills (Ott, 2010), which were curriculum-specific (Cruse & Twing, 2000). TEAMS measured a slightly higher level of skills (McNeil et al., 2008) and tested students in odd grades one, three, five, seven, nine, and eleven (Champion, 2007; Haney, 2000; Ott, 2010). Students who failed the test were

mandated to receive remediation and then be retested (Ott, 2010). Eleventh graders had

to pass the test to graduate (Champion, 2007; Ott, 2010). The ninth-grade class of 1987

was the first to be required to pass exit tests to graduate (McNeil et al., 2008).

The shift from TABS to TEAMS was largely attributable to the national context

in the 1980s. Under the influence of the ANAR report, business leaders in Texas pushed

school reforms (Vasquez Heilig et al., 2012) and the state increased its involvement in

education policies. H. Ross Perot, a prominent Dallas businessman, and his allies played

an essential role in creating accountability and testing (Carnoy & Loeb, 2002). Based on

recommendations from a 1984 commission on school reform led by Perot, House Bill 72

(HB72) introduced many education reforms that are still relevant today (Mumby, 2010;

Stutz, 2014). HB72 changed the administration and oversights of the public education

system. Relevant testing policies included high school graduation tests and the no-pass,

no-play rule for high school student athletes (Stutz, 2014). Texas was the first state

enacting the no-pass/no-play rule into law (which was later amended); many other states

followed suit (Mumby, 2010; Stutz, 2014). HB72 also included "a provision that called

for all educators to pass competency test to keep up their job," which incurred strong

opposition to the reelection of the then-governor (Stutz, 2014). Regardless, Texas

became one of the earliest states to develop statewide testing systems in 1987 and

implement minimum competency tests for high school graduation (Vasquez Heilig et al.,

2012). At that time, testing was mainly used to check whether or not educational

programs were effective, and this enabled the public to get a sense of what and how

schools were teaching (A. Smisko, personal communication, March 17, 2015).

**TAAS**.  The 1991 Texas Legislature increased assessment rigor and created the third test, the Texas Assessment of Academic Skills (TAAS).  This pre-NCLB test was in place until 2002 (Champion, 2007; Haney, 2000, 2001; Ott, 2010).  TAAS was tied to the new state curriculum of Essential Elements and administered to students in grades three, five, seven, nine, and eleven in 1990 (Champion, 2007; Ott, 2010).  When TAAS was moved from fall to spring in 1993, third through eighth and tenth graders were tested in reading and mathematics; fourth through eighth and tenth graders were tested in writing (Ott, 2010).  After 1993, students in grades four, eight, and ten consistently took reading, mathematics, and writing tests in the spring (Haney, 2000).  The eleventh-grade students had to pass TAAS exit tests to graduate (McNeil et al., 2008; Ott, 2010).  While exit tests were state mandated, alternative routes existed for failing students.  In the late 1990s, students failing the eleventh grade TAAS test could take EOC exams in selective subjects to graduate (Haney, 2000; Ott, 2010).  In the late 1990s, testing scores were also used for grade-level promotion.  In 1999, the 76th Texas Legislature introduced the Student Success Initiative (SSI) that required third, fifth, and eight graders to pass a test in order to advance to the next grade.

TAAS shifted focus on testing from minimum skills to academic skills (Cruse & Twing, 2000; TEA, 1997).  The assessment was not only used to assess public school students' "higher-order thinking skills and problem solving ability" (TEA, 1997, p.1), but also linked to important educational decisions.  In 1993, as a result of the Texas legislature, TAAS "became the linchpin of the accountability system, the first to have consequences for children, schools, and districts" (McNeil et al., 2008, p.3).  Schools and

districts were held accountable for student learning on "academic excellent indicators," such as testing data and dropout and attendance rates (TEA, 1997, p.159).  In 1994, the then-superintendent of Houston Independent School district, Rod Paige, based "principals' performance on each school's test scores…This change formalized the school as the primary unit of analysis in the accountability system" (McNeil et al., 2008, p.4).  Student testing data also had to be disaggregated by race/ethnicity and socioeconomic status and a school was rated unacceptable if all subgroups did not meet the standards (Haney, 2000).

Increased accountability during the TAAS period was largely politically-driven. This had a lot to do with the beliefs of the then Texas Education Agency (TEA) commissioner, the legislative requirements, and governors' decisions.  From 1991 to 1995, Skip Meno was TEA commissioner.  He believed it was not good enough for public schools to teach middle-class students; rather schools should be responsible for teaching all kids, regardless of color and SES, and it was important to hold the statewide system accountable (A. Smisko, personal communication, March 17, 2015).  As a result of the Senate Bill 7 (SB7) mandate that was signed into law by the then governor Ann Richards in 1993, Texas created an education system that evaluated both districts and campuses (Vasquez Heilig et al., 2012).  Under the leadership of Governor Richards, Texas enacted its first accountability system in 1994 (Vasquez Heilig et al., 2012).  This system functioned as "an information forum that utilized test scores and other measures of student progress to determine whether school districts should remain accredited by the state" (Vasquez Heilig et al., 2012, p.563).  At that time, the Texas accountability system was "undergirded by the Public Education Information Management System (PEIMS)

data collection system, a state-mandated curriculum, and a statewide standardized test to measure student proficiency in core subjects" (Vasquez Heilig et al., 2012, p.564). SB7 signaled a shift of the Texas accountability system "from a diagnostic to a performance-based system" (Vasquez Heilig et al., 2012, p.563). Between 1995 and 1999 when Governor George W. Bush was serving his first term, the Texas test-based accountability system raised its stakes by moving beyond districts and "applying a variety of sanctions on teachers, principals and schools" (Vasquez Heilig et al., 2012, p.564). The strengthened system was largely associated with promoting Governor Bush's reading initiative (A. Smisko, personal communication, March 17, 2015). The implementation of the initiative took the form of Senate Bill 1 of 1995, which required districts to achieve certain reading goals (Denton, n.d.).

**TAKS**. Due to the mandate of the 76th Texas Legislature in 1999, TEA adopted the new state assessment program of the Texas Assessment of Knowledge and Skills (TAKS), administered the field test in 2002, and then tested statewide in spring 2003 (TEA Curriculum Division, personal communication, March 13, 2015). TAKS aligned with the new state-mandated curriculum–the Texas Essential Knowledge and Skills (TEKS)–that went through careful revisions and public reviews (D. Nance, personal communication, Feb 4, 2015). One group of external actors that sought to raise academic standards were business representatives who wanted students to be prepared to compete in the market (D. Nance, personal communication, Feb 4, 2015). Also, as a response to the federal NCLB requirement, TAKS was in place until 2010, focusing more on "higher level thinking skills, measured by students' abilities to analyze information, develop

sound conclusions and apply principles to problem-solving" (D. Nance, personal communication, Feb 4, 2015).

TEKS essentially mandated that state assessments should be aligned with state curriculum (D. Nance, personal communication, Feb 4, 2015). TEKS and TAKS were developed during the late periods of TAAS (TEA Curriculum Division, personal communication, March 13, 2015). Prior to TEKS, the state curriculum–the aforementioned Essential Elements–held expectations for teachers rather than students (TEA Curriculum Division, personal communication, March 13, 2015). Commissioner Meno ushered in the perspective shift of focusing more on students when implementing curriculum during the TAAS period (D. Nance, personal communication, Feb 4, 2015). Initiated by the State Board of Education (SBOE), statewide selected committees of educators, parents, business and industry representatives, and employers developed TEKS (D. Nance, personal communication, Feb 4, 2015; TEA Curriculum Division, personal communication, March 13, 2015). SBOE adopted TEKS for all subjects to be effective on September 1, 1998, when Meno's successor Michael Moses (1995-1999) was TEA commissioner (TEA Curriculum Division, personal communication, March 13, 2015). To align with TEKS's higher curriculum standards, TEA revised TAAS to measure what students should learn and adopted TAKS (TEA Curriculum Division, personal communication, March 13, 2015).

TAKS was more rigorous than previous tests, as it assessed more subjects and grade levels (Champion, 2007; Ott, 2010). The tested subjects in grades three through eleven included reading, writing, mathematics, social studies, and science. Grade four

writing and grade eight social studies were state-administered, but not required by NCLB. High school students had to pass TAKS exit testing in English language arts, social studies, math, and science to graduate; they could take the tests up to five times during their junior and senior years. Although new students enrolling after January 1 could substitute reading and math with the SAT or ACT, they still had to take TAKS exit exams in science and social studies.

**STAAR.** The State of Texas Assessments of Academic Readiness (STAAR) test, which began in spring 2012, was in accordance with legislative requirements. According to TEA public data, Senate Bill (SB) 1031 of 2007 and House Bill 3 (HB3) of 2009 led to the new generation of testing, developed through TEA's collaboration with the Texas Higher Education Coordination Board and Texas educators. The test format of STAAR in grades three through eight is comparable to TAKS. The new tests include the general test and four variations of tests for special groups of students (Guthrie, 2011). STAAR continues to test the content standards of the state curriculum TEKS, but the new tests more deeply assess TEKS than TAKS and emphasize the skills that are needed for the next grade level or course (ultimately for college and career). STAAR tests "fewer skills," but it evaluates these skills more broadly and in-depth (Guthrie, 2011; Young, 2012), emphasizing more critical thinking and aiming for postsecondary readiness (Young, 2012). The accountability system for 2013 and beyond was specified to include multiple academic measures, including student performance on grades three through eight and EOC assessments, high school dropout rates, and graduation rates. Under STAAR, students in grades five and eight had to also pass the STAAR reading and mathematics to

be advanced to the next grade level, which was 81$^{st}$ Texas Legislature's modification in 2009 to the 1999 SSI grade advancement requirements (TEA, 2014a).

Increasing the rigor of Texas accountability, STAAR differs from TAKS in several ways. Based on TEA information, the new tests assess academic skill at a greater depth and level of cognitive complexity. STAAR also increases the performance standards for all grades and subjects and the total number of test items for most grades and subjects. Additionally, STAAR implemented EOC exams instead of exit tests that had been in place for over three decades and would add two voluntary EOC assessments of English and algebra in 2016 (TEA, 2014b). Starting in the 2011-12 school year, students took three EOC exams in four core subjects: English, mathematics, science, and social studies. The statutes originally proposed that EOC assessments counted for 15 percent of a student's final course grade (Guthrie, 2011). Partly due to pressure and resistance from parents and schools across the state, HB5 of 2013 repealed the 15 percent grading requirement, leaving the decision of EOC weight to local school districts. HB5 also expanded accountability measures by adding financial accountability and the evaluation of community and student engagement (CASE) to the Texas educational accountability mechanism. This CASE policy is the focus of this dissertation.

**Summary**. Over the past three decades, the state of Texas has tightened and increased its control over accountability of districts and schools. Although local schools had certain flexibility in determining how to achieve the goals of the reading initiative put in place by Governor George W. Bush and earlier graduation exit tests gained alternative routes in previous years, testing revisions are top-down political decisions and the

accountability system remains centralized.  Texas tests more subjects that are not required by NCLB and increases high school graduation requirements.  The state also increases curricular and assessment standards and shifts towards college and work readiness, largely a response to the business community.  Partly owing to public inputs, the 2013 statutory changes reduced the number of EOC tests, returned the use of EOC assessments in students' grades to local school districts, and added a local assessment element to the accountability mechanism.  Local control remained minimal in the process of educational decision-making.

**The Impact of Test-Based Accountability**

In the preceding sections, I reviewed the trend of increased federal control in education before and after NCLB and its recent change, as well as strong state control in initiating several generations of high-stakes testing and its new shift in Texas.  In this section, the review of the impact of high-stakes testing on students, schools, and teachers will shed light on how nation-wide resistance to the top-down high-stakes testing accountability system came about, as documented in chapter one.  Findings in extant literature on the impacts of high-stakes testing also lay the groundwork for understanding calls for accountability reforms that are to be examined in the following segment.  The literature regarding how high-stakes testing (before and after NCLB) affected students, teachers, and schools is organized by five themes: student achievement, student attainment, achievement gap, student placement, and impact on schools and teachers.

**Impact on students: academic achievement.**  Standardized and high-stakes testing are implemented in the hopes of increasing academic achievement for all students

but, as this review demonstrates, they have not yielded the hoped-for gains.  Some

researchers reported negative and null impacts, while others presented mixed findings.

  ***Negative or null impacts***.  Many previous studies showed the negative effect or

zero effect of accountability policies driven by high-stakes tests (Amrein & Berliner,

2002; Jacob, 2005; Lee & Wong, 2004; Marchant & Paulson, 2005; McCarty, 2009;

Nichols, Glass, & Berliner, 2006; Ravitch, 2010).  Some research used national tests to

examine the effectiveness of high-stakes testing policies in the states that attached serious

sanctions to their testing system.  Before NCLB, research found no evidence for

increased student learning in 18 states after the implementation of high-stakes testing

policies (Amrein & Berliner, 2002).  Still, literature after NCLB did not demonstrate a

relationship between testing pressure and achievement for fourth and eighth grade

National Assessment of Educational Progress (NAEP) math and reading in 25 states

(Nichols, Glass, & Berliner, 2006).  Additionally, exploring the relationship between

statewide testing and student performance, one study reported no evidence of Native

American students' achievement gains (McCarty, 2009).  Further, researchers reported no

positive achievement effect for the lowest-performing students, and reduced performance

"for racial and ethnic minorities and women in some subjects by activating stereotype

threat" (Holme et al., 2010, p.16).  Many studies found English language learners (ELLs)

were more likely to experience more negative impacts due to language barriers.  In

California, only 19 percent of ELLs passed the state exit exam after two attempts

(Rumberger & Gándara, 2004) and there was little evidence of achievement improvement

for Arizona ELL students (Wright & Choi, 2006).

43

*Mixed findings.* A significant amount of research has produced mixed findings when examining state testing impacts before and after NCLB (Carnoy, Loeb, & Smith, 2001; Fuller, Wright, Gesicki, & Kang, 2007; Hanushek & Raymond, 2003, 2004; Linton & Kester, 2003; Neill, 2001; Vasquez Heilig & Darling-Hammond, 2008). Fuller, Wright, Gesicki, and Kang (2007) found fourth graders' earlier test score increase faded after the enactment of NCLB; alternatively, the math achievement increase persisted but the growth was at a slower rate than before. Linton and Kester (2003) revealed while Texas's TAAS mean scores dramatically increased for all students between 1996 and 2000, the state's NAEP score growth applied only to Hispanic students. Likewise, Vasquez Heilig and Darling-Hammond (2008) showed the sharp increase in state-reported TAAS scores as a result of excluding low-achieving students was not reflected in the low-stakes Stanford Achievement Test-Ninth Edition.

Varying effects on distinct student groups, grades, or subjects also reflected the mixed findings of high-stakes testing (Center on Education Policy, 2010; Gándara & Baca, 2008; Jennings & Beveridge, 2009; Lauen & Gaddis, 2012; Riderick, Bryk, Jacob, Easton & Allensworth, 1999; Rumberger & Gándara, 2004; Shuster, 2012; Vasquez Heilig, 2011; Wright & Choi, 2006; Wright & Pu, 2005). For instance, Lauen and Gaddis (2012) found accountability sanctions had positive effects for minority and disadvantaged students and larger positive effects for the lowest-performing schools than schools near the passing margin, but produced adverse impacts for all students in math. Using federal data, Shuster (2012) showed negative effects of exit exams for students in the bottom-two quartiles, and slight positive effects for the top quartile students.

44

**Impact on students: academic attainment.**  Many states have imposed
additional high school graduation requirement tests (exit exams) that were not required
by NCLB.  There are concerns that exit tests have exerted unintended deleterious effects
on student attainment, driven significant numbers of disadvantaged students out of school,
and lowered the quality of public high schools (Darling-Hammond, 1991, 2004; Haney,
2000; Hong & Youngs, 2008; McNeil, 2005; Neil, 2003; Vasquez Heilig & Darling-
Hammond, 2008).  This section reviews relevant literature and identifies the positive, null,
and negative impacts of high-stakes graduation requirements on academic attainment.

*Positive or null impacts.*  A small number of studies report either positive or zero
impacts of high-stakes testing on students' academic attainment.  Carnoy, Loeb, and
Smith (2001) found that Texas high school progression and graduation rates had a slight
increase in the 1990s and in early 2000, and the larger pass rates of tenth graders on
Texas TAAS echoed the larger declines in dropout rates.  The following year, these
researchers reported similar results after evaluating the strength of accountability in 50
states between 1996 and 2000.  This study suggested that high-accountability states did
not have significantly higher retention levels or lower high school completion rates
(Carnoy & Loeb, 2002).  Using federal data, Shuster (2012) reexamined the impact of
high school exit exams, showing that exit tests basically had no effect on school
completion and did not affect GED seeking and acquisition.  Additionally, situating their
work in Florida and Texas and using the 1998-2000 October Current Population Surveys,
Warren and Jenkins (2005) found no evidence that state high school exit exams alone
accounted for higher dropout rates or greater inequalities in dropout rates.  However,

Vasquez Heilig (2011) attributes Warren and Jenkins' findings to the oversampling of white students and improper categorization of minority populations.

*Negative impacts*.  According to literature, state-mandated exit testing prior to NCLB was associated with lower school completion rates and higher rates of General Educational Development (GED) test taking; this association grew stronger when states had more minority and poor students (Warren, Jenkins, & Kulick, 2006).  A comprehensive review reinforced this observation, indicating that high school exit exam policies were associated with increased drop-out rates, especially for low-achieving and poor urban students (Holme et al., 2010).  Others found that national high school graduation rates declined (Miao & Haney, 2004) and high school exams were associated with decreased graduation rates even after controlling for students' demographics (Marchant & Paulson, 2005).  Exit exams also affected the severity of the drop-out issue (McNeil et al., 2008). ELLs were more likely to drop out and less likely to graduate when they were tested in English, a language they have not mastered (Menken, 2010; Orfield, Losen, Wald, & Swanson, 2004).

Numerous studies focus on the impacts of exit tests on non-white and minority students.  The number of minority drop-outs in some districts was alarming (Losen, Orfield, & Balfanz, 2006; Orfield, Losen, Wald, & Swanson, 2004).  Low-achieving students had an increased probability of dropping out (Jacob, 2001); in some states those students were pushed out of schools (Orfield, Losen, Wald, & Swanson, 2004) or encouraged to leave (Vasquez Heilig & Darling-Hammond, 2008).  Poor students, ELLs, African Americans, and Latinos were more likely to be pushed out of schools in Texas

46

(McNeil et al., 2008) and low-income urban students from Massachusetts had a higher likelihood of dropping out despite their on-time grade promotions (Papay, Murnane, & Willett, 2010).  A large number of Latinos did not receive high school diplomas owing to the failure of exit exams (Valenzuela, 2004).  Some literature also demonstrated the negative psychological effects of exit tests on students (Holme et. al., 2010; Vasquez Heilig, 2011).

**Impact on students: achievement gap.**  A laudable goal of NCLB is making schools accountable for learning among all student groups.  A significant amount of research has examined whether or not the achievement gaps between white students and minority students, English learners and non-English learners are narrowed or closed.  Some of the literature analyzes the achievement gap on NAEP or other low stakes state standards, primarily testing the validity of state-reported improvement.  As the following literature shows, it seems that the achievement gap after NCLB has not been bridged.

*The gap between whites and other students*.  This gap is usually studied and does not seem to change much.  Researchers argue that certain student groups, such as African American students and Latinos/Latinas are already low-performing and would be more disadvantaged in high-stakes testing (Brennan, Kim, Wenz-Gross, & Siperstein, 2001).  For instance, Hanushek and Raymond's (2003) analysis of NAEP data show that the black/Latino-white achievement gap did not close but increased after NCLB.  One year later, the same authors (2004) report that the black-white gap was not narrowed.  Nearly a decade later, Reardon, Greenberg, Kalogrides, Shores, and Valentino (2013) find no positive evidence for the substantial narrowing of racial achievement gaps.  Using NEAP

47

data between 2000 and 2007, Braun, Chapman, and Vezzu (2010) indicate that the black-white achievement gap was large and NCLB had a slight impact on reducing the gap between the two groups. Similarly, Ravitch (2010) claims that the gap between black and white students was narrower before NCLB than in the years afterward. Interestingly, analyzing NAEP grade data between 1992 and 2006, Fuller, Wright, Gesicki, and Kang (2007) note that the narrowing achievement gaps in the 1990s had basically disappeared in the post-NCLB era. A most recent report produces another complex picture of state variation in narrowing the achievement gap (Reardon et al., 2013). These findings show that NCLB overall had no effect on the achievement gap. The researchers have found, however, some effects that vary by state context. White-black and white-Hispanic achievement gaps have narrowed in states with more subgroup accountability pressure, more across-school segregation, and large existing gaps, but appear to have increased in states that face less pressure, less segregation, and have small existing gaps (Reardon et al., 2013).

The achievement gap change is inconsistent on state and national testing. Whereas a smaller and decreasing gap between whites and African Americans existed on Texas TAAS, that gap was larger and increased on NAEP results (Haney, 2000; S.P. Klein, Hamilton, McCaffrey, & Stecher, 2000) or saw no change on NAEP (Linton & Kester, 2003). A small gap in state testing between Texas whites and Latinos became larger in terms of NAEP data (McCombs, Kirby, Barney, Darilek, & Magee, 2005). After comparing TAAS to SAT-9, Vasquez Heilig and Darling-Hammond (2008) suggest that although the achievement gap between white and minorities appeared to be

substantially closed on the high-stakes TAAS, it was slightly reduced on the low-stakes SAT-9.

*The gap between ELLs and non-ELLs.* Both early and recent literature do not find that the achievement gap between ELLs was bridged. ELLs generally scored lower than non-ELLs on reading, science, and math in terms individual items and total test scores (Abedi, 2002). In California, the achievement disparity between ELLs and non-ELLs was not only large but also increased for students in higher grades (Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Rumberger & Gándara, 2004). Researchers report serious gaps between ELLs and their counterparts in Arizona (Wright & Pu, 2005). Recent evidence shows that the gap has noticeably increased on Texas TAAS math and increased more on TAAS reading (Vasquez Heilig & Darling-Hammond, 2008) and the gap has increased in the TAKS era compared to the previous TAAS era (Vasquez Heilig, 2011); large gaps have cut across content areas, grades, and states (Kim & Herman, 2009); and the gap in high school reading proficiency is very large based on an examination of 35 states' achievement data (Center on Education Policy, 2010).

**Impact on schools and teachers**. Research shows that before and after NCLB, schools would be probably incentivized to inflate scores by manipulating the testing pool (Darling-Hammond, 2004, 2007; Nichols & Berliner, 2007; Valenzuela, Fuller, & Vasquez Heilig, 2006) or teaching to the test (McNeil & Valenzuela, 2001; Ravitch, 2010; Sloan, 2007; Vasquez Heilig, 2011). This section explores how schools and educators respond under the pressure of states' test-based accountability and NCLB in deciding who to include in the testing pool and what to teach in the classroom. In general, high-

49

stakes testing drives schools to adopt various methods of student placement, administer

test preparation, and narrow curriculum.

***Student placement.*** Extant research documents the variety of methods schools

adopt to manipulate the testing pool in order to raise test scores and accountability ratings.

These strategies include test exemption (Jennings & Beveridge, 2009; Vasquez Heilig &

Darling-Hammond, 2008), exclusion (Amrein-Beardsley & Berliner, 2003; Booher-

Jennings, 2005; Darling-Hammond, 2007; Hong & Youngs, 2008; Nichols & Berliner,

2007; Vasquez Heilig & Darling-Hammond, 2008), special education placement

(Allilngton & McGill-Franzen, 1992; Booher-Jennings, 2005; Figlio & Getzer, 2002;

Vasquez Heilig & Darling-Hammond, 2008), grade retention (Carnoy, Loeb, & Smith,

2001; Nichols, Glass, & Berliner, 2006; Vasquez Heilig & Darling-Hammond, 2008),

and encouragement to transfer to a GED program (Haney, 2000). In particular, low

achieving students and/or minority groups are more likely to be displaced.  Jennings and

Beveridge (2009) report that African American and Latinos are more likely to be

exempted.  Booher-Jennings (2005) note that schools triaged students based on academic

performance.  Darling-Hammond (2007) finds that schools excluded low-scoring students

from testing in order to achieve their test score goals.  Using national data, Nichols and

Berliner (2007) shows how most teachers focused on "bubble kids" and viewed low-

performing students as a burden rather than an opportunity to foster educational

improvement.  In some state, schools excluded elementary students through special

education and language exemption (Vasquez Heilig & Darling-Hammond, 2008).

Retention is also practiced in some schools within some districts to prevent students from

50

promoting to the testing grade, which would artificially inflate testing scores. Carnoy, Loeb, and Smith (2001) suggest that African American and Latino students are more likely to be retained. Literature shows that whereas some schools chose to increase the retention of ELLs (Vasquez Heilig, 2011), others tended to retain ninth graders so that low-performing kids would not take the tenth-grade exit tests (Vasquez Heilig & Darling-Hammond, 2008).

*Curriculum and instruction.* Finally, some researchers find that the unintended consequences resulting from the testing-driven accountability strategies are test preparation and narrowed curriculum (Darling-Hammond, 2007; Hong & Youngs, 2008; McNeil & Valenzuela, 2001; Ravitch, 2010; Valencia & Bernal, 2000; Vasquez Heilig, 2011). In some Texas schools, poor and minority youths were taught the lowest level of information and skill and test preparation activities (McNeil & Valenzuela, 2001). Examining the impact of high school exit testing on English language learners, Vasquez Heilig (2011) found that in some schools of Texas many courses were TAKS focused and there were only a few days for intensive test training. Literature also reports that ELLs in Arizona received testing-focused education that did not meet their linguistic and academic needs (Wright & Choi, 2006).

**Summary and analysis.** My literature review demonstrates that state-mandated testing before NCLB and high-stakes testing based accountability in the post-NCLB era have yielded few, if any, benefits, and that as a whole it has produced null or negative results. These findings echo some recent publications. The ineffective NCLB system is seen in American students' performance on international tests. Darling-Hammond (2013)

finds that test-based accountability did not produce gains from 2000 to 2012 on the Program for International Student Assessment.  Discussing the reauthorization of ESES, Welner and Mathis (2015) note a broad consensus among researchers, who have found that NCLB-style, high-stakes, testing-focused systems are "at best ineffective and at worse counterproductive" (Welner & Mathis, 2015, p.1).

**Alternatives to High-Stakes Testing Accountability**

The theory underlying NCLB and Texas-style accountability is that when we hold students and school accountable through testing, scores will automatically improve education for all groups of students because educators, schools, and students will work harder (Vasquez Heilig, 2011).  The motivation to try harder stems from the pressure that disaggregating testing scores by demographics define student success (Hanushek, & Raymond, 2003).  When both federal and state governments increase assessment standards without increasing resources to local districts and schools, local entities are constrained to achieve the federal and state accountability goal.  Under the constraints of resources and the pressure of being punished, schools and districts might utilize strategies to boost test scores without improving all students' academic achievement, as the preceding literature review shows.  As Darling-Hammond (2007) noted, the current policy is unable to fulfill its intended goals if the problems of unequal educational resources are not addressed.

Nonetheless, the current high-stakes testing policy has raised the academic bar, created national awareness for disadvantaged students, and pushed schools to ensure those students' success.  Unintended and negative consequences together with extensive

resistance to high-stakes testing have drawn attention from state governments and pushed some states to revise the existing test-based accountability system. As documented in Chapter 1, by 2014 legislators from some states have decreased the number of required tests (Guisbond, 2014) and more are making statutory changes to mitigate their high-stakes testing system in 2015. The Rhode Island legislature voted to stop using high stakes-testing for three years (Hagopian, 2014). A new assessment system in Connecticut goes beyond test scores (Megan, 2014). Pennsylvania tried to end the use of the high-stakes testing for high school graduation (Ravitch, 2015). The Mississippi House no longer requires the testing of four subjects for graduation (Amy, 2015) and Georgia removed graduation test requirement (Redmon, 2015).

Some researchers propose alternatives to the accountability assessment system. Focusing on Texas, Valenzuela, Sun, Germain, and Barnes (2015) suggest in their policy brief that the state accountability system should move toward an informational framework characterized by non-discretionary measures and discretionary measures. Standardized test score is one type of non-discretionary measure and indicators reflecting local circumstances or values are discretionary measures (Valenzuela et al., 2015). The informational model seems to be a balance of state and local control. Another model is community-based accountability (CBA), under which a community is "empowered to be accountable to themselves and to the nation…which communities democratically set achievement and outcomes goals" (Vasquez Heilig, Khalifa, & Tillman, 2013, p.533). CBA is an ideal of local control where state and federal control have been largely reduced. Specifically, the roles of the state and federal government "would be to

calculate baselines, growth, and yearly ratings (recognized, low-performing, etc.) for a set of goals that the community selected in a democratic process" (Vasquez Heilig et al., 2014, p. 877).

In practice, there emerges new assessment of non-academic measures and more local control in a few states. As previously mentioned in Chapter 1, California and Texas adjusted their accountability systems to incorporate non-test measures and experimented with local models in 2013. Using a state-approved LCAP template, local educational agencies in California design their budget plans and develop annual goals and actions to meet eight state priorities and school districts report progress in seven areas in addition to student achievement (Kerchner, 2014). The Texas Legislature added alternative non-testing measures–community and student engagement (CASE)–to its test-based accountability system in 2013. This new statute calls for local districts and schools, through a local committee, to adopt their own criteria to self-evaluate performance in eight required academic programs and one compliance category. This study focuses on the CASE policy, which was immediately implemented in the 2013-14 academic year. Little empirical knowledge exists about the implementation process and the effect of this new policy. The study fills in the gap to generate timely empirical evidence to inform Texas and other states of revising their education accountability systems and responding to the ESSA requirements of keeping public schools accountable in non-academic measures as well as academic domains.

**Theoretical Framework**

To guide this dissertation and evaluate its findings in chapters 4 and 5, I use the theory of sensemaking (Weick, 1995), the political model (Malen, 2006), and the role of internal institutional conditions (McLaughlin, 1987). In this section, I first discuss sensemaking theory by providing a brief view of its content and discussing its application to district leaders' perceptions of the CASE policy. I then introduce the political framework and discuss the application to the process of determining CASE rating criteria in the local districts of my case studies. Finally, I present lessons from policy implementation and talk about the application to the implementation process and results associated with my study.

**Sensemaking.** Theories of sensemaking are grounded in the microsociological traditions that emphasize how social interactions affect the interpretation of events (Weick, 1995). One finding about sensemaking in literature is the effect of contexts on people's sensemaking or perceptions. Previous research on math and language arts standards in several states has found that different interpretations of the same policy message can result from people's cognitive understanding and contexts (Cohen & Barnes, 1993; Spillane, 1996, 1998). A study on how teachers responded to accountability reform also suggests the significance of settings under which policies work (Coburn & Stein, 2006).

Following this line of logic, how district leaders perceive CASE might differ by varying district contexts. While the CASE policy stems from the state legislative requirement, it is locally interpreted, defined, and implemented. It is reasonable to imagine that people from diverse types of districts perceive the same policy differently.

It is possible that district leaders from urban districts that serve diverse students and those from rural districts that have more local control perceive CASE differently from their counterparts from other types of districts, such as suburban and non-metropolitan districts. It can also be expected that district leaders' perceptions are associated with their districts' academic contexts when confronted with a new policy initiative like CASE. That is, district leaders from districts of various community types and with different accountability ratings may perceive CASE differently. Overall, the theory of sensemaking is particularly useful in interpreting the findings of first research question: How do public school district leaders perceive the CASE policy? It allows for an investigation of potential variations in perceptions across district types and district academic status.

**The Political Framework.** Malen's (2006) political framework is useful in understanding how district implementers engage local community stakeholders into the process of determining CASE rating criteria. The political model posits that "policy implementation is a dynamic political process that affects and reflects the relative power of diverse actors and the institutional and environmental forces that condition the play of power" (Malen, 2006, p. 85). In particular, this framework pays attention to "the realities and complexities of 'politics' in whatever form those dynamics may take" (Malen, 2006, p. 85). In this sense, the power dynamics between diverse actors during the process of policy implementation is likely to differ in different contexts.

In terms of my study, the extent that district and campus staff, teachers, parents, and community members are involved in developing CASE rating criteria might also

vary from district to district. While "actors at all levels of the system can influence policy implementation" (Malen, 2006, p.86), the weight of diverse actors may vary in different contexts. As CASE primarily evaluates campus' performance, district implementers might involve campus staff more than other actors, such as parents or community members. Depending on districts' existing initiatives and resources, and how they interpret the definition of community engagement and student engagement, the involvement of teachers, parents, business representatives, and community members might differ across districts, with some having strong voices and others being silent in the policy implementation process.

Another significant component of the political framework is the intersection of actor resources and policy currencies (Malen, 2006). This framework suggests that the resources actors can capitalize upon to exercise influence determine their impact on policy developments. The resources can take the forms of "financial, informational, social and cultural capital" (Malen, 2006, p. 88). Applying this to the policy implementation arena, actors' resources also influence their participation in policy implementation. For my study, parents, business, and community members, who possess expertise and knowledge in the evaluated areas of CASE or who are available to participate, are more likely to be invited to weigh in the CASE rating criteria development than their counterparts. This suggests that low-income parents or new immigrant community members might be less likely to be involved in the CASE policy implementation process.

Overall, the political perspective is particularly useful in interpreting the findings of the second research question: How have public school districts implemented the CASE policy? It is interesting to explore the different inclusion patterns of district and campus staff, teachers, parents, business representatives, and community members in the CASE process. Given that CASE is locally interpreted and evaluated, it is possible that the districts participating in my qualitative case studies demonstrate varying implementation patterns.

**Internal Institutional Conditions.** To understand the implementation and impact of the CASE policy, it is also important to understand the influence of "internal institutional conditions" on policy implementation (McLaughlin, 1987, p.173). Whether or not a context is stable or there are competing priorities will significantly influence implementers' actions (Yin, 1981, as cited by McLaughlin, 1987). For instance, when a teachers' strike diverted a great deal of resources and attention, a school improvement program in California was ignored in a large school district of this state (McLaughlin, 1987). Applying the first lesson to my study of CASE, it is possible that districts that are unstable, such as experiences budgets cut or leadership change, are likely to invest less in developing their CASE rating criteria. Likewise, districts with existing priorities or many initiatives are less likely to engage few stakeholders into the process of determining CASE rating criteria.

McLaughlin also points out the importance of having a balance of pressure and support in effective implementation of a policy (McLaughlin, 1987; Elmore & McLaughlin, 1982; McLaughlin & Pfeifer, 1988). For instance, in the absence of district

support, a policy can mandate districts to "establish a parental involvement mechanism," but "cannot require them to welcome parents and facilitate their participation" (McLaughlin, 1987, p.173). For my study, if districts do not hold positive perceptions of or support an initiative such as CASE, their implementation is likely to be more compliance-oriented. This is an imbalance of pressure and support. Conversely, when there is a balance, that is districts perceive CASE positively and have district or community support in place, they are likely to effectively implement CASE and make improvements on the evaluated areas.

# CHAPTER 3 METHODOLOGY

This exploratory study seeks to understand how school district administrators perceive the CASE policy, how they have implemented the policy, and how they perceived the CASE impact. The policy under study is new and has hardly been studied. Creswell (2003) suggests that an exploratory study is appropriate if research subjects have been the topic of little prior research. In this chapter, I lay out a specific plan to fulfill the purpose of this study. I provide the research questions and related methods for studying them in the table below (Table 1). To understand district leaders' perceptions and their implementation of the new policy, I employ a parallel mixed-methods design (Teddie & Tashakkori, 2009), concurrently collecting both survey data and interview data. I specify methods of data collection and analysis for the survey study and qualitative case studies, respectively, and discuss limitations and strengths of this study.

**Table 1:**

*Research Questions and Methodological Approach*

| Research Questions | Related Methods and Analysis |
|---|---|
| 1. How do public school district leaders perceive the CASE policy? | quantitative survey data: descriptive and inferential statistics |
| 2. How have public school districts implemented the CASE policy?<br>    a. Who was involved in determining the rating criteria?<br>    b. What was the process for determining rating criteria? | qualitative survey data: cross-case analysis<br> interview data: cross-case analysis |
| 3. How do public school district leaders perceive the impact of the CASE policy? | qualitative survey data: descriptive statistics<br>interview data: cross-case comparison |

**The Survey Study**

Surveys are "information methods used to describe, compare, or explain individual and societal knowledge, feelings, values, preferences, and behaviors" (Fink, 2013, p.2). They can be used for policy needs and research, to plan or evaluate programs (Fink, 2013). Because I am interested in understanding how the CASE policy is perceived and implemented in the state of Texas, the survey method is most appropriate for answering my research questions about district leaders' perceptions of the CASE policy (Research Question #1) and its impact (Research Question #2), and their implementation of the policy (Research Question #3). The survey data collection method is self-administered online questionnaires. Based on Fink (2013), online surveys are appropriate if "the survey needs wide geographic coverage" and the aim is to "reach a very large number of people" (p.11). This is the case for my study, which focuses on districts statewide and tries to survey a large number of district leaders. In what follows, I discuss survey development, the sampling, data collection, data entering, and analysis.

**Survey development**. For this study, I surveyed district leaders' perceptions of the CASE policy and its implementation through online questionnaires. Likert-scale questions on perceptions constitute the first part of the survey and descriptive questions regarding what district leaders did to implement the CASE the policy are the second part of my survey. The third section consists of open-ended questions. Because I am interested in the ways in which different districts in different contexts perceive the policy, the final part of this survey is centered on demographic questions. As I noted in the preceding chapter, districts in different contexts may differently perceive and implement

the CASE policy. For instance, rural districts may lack the staff to act independently and therefore may find it hard to implement the policy; urban districts might lack resources, feel besieged by various demands, and perceive the CASE policy a burden.

Given that I did not identify an existing survey instrument to replicate, I used my literature review, as well as existing qualitative interview data on the CASE policy from a pilot study conducted in a previous semester, to draft a new survey questionnaire. According to Mauceri (2004), qualitative data could be effectively used to improve the quality of a survey instrument for new studies. Data from the pilot qualitative interview gives me an idea of how some district leaders perceived the new CASE policy and what they did to evaluate the new accountability measures of (1) fine arts, (2) wellness and physical education, (3) community and parental involvement, (4) twenty-first-century workforce development programs, (5) second language acquisition programs, (6) digital learning environment, (7) dropout prevention strategies, (8) educational programs for gifted and talented students. Qualitative data enabled me to develop relevant questions about attitudes and behaviors relative to the CASE policy. To construct questions, I also relied upon literature review, which is a common and important way to supplement questions that might be omitted and missing in qualitative data. The review of attitudinal survey studies on education issues and policies is relatively extensive (Albirini, 2006; Avramidis, Bayliss, & Burden, 2000; Branton, 2007; B. Cavas, P. Cavas, Karaoglan, & Kisla, 2009; Y.F. Chang, 2008; Cook, Semmel, & Gerber, 1999; Guskey, 1988; Mehta, & Kellert, 1998; Rose & Gallup, 2007; Shade & Stewart, 2001; Tait & Purdie, 2000).

These studies helped me double check the phrasing of survey questions and modify

demographic questions that relate to for people's attitudes and behaviors.

As previously stated, the survey consists of four parts with a total of twenty-nine

survey questions (Appendix B). The first section contains ten questions asking about

school district leaders' perceptions of the CASE policy, which answers the first research

question. As in previous studies, the respondents were "asked to indicate the extent of

their agreement with each statement" by selecting among response choices (Avramidis,

Bayliss, & Burden, 2000, p.196). To identify variations in the findings, I chose to use

Likert-type scales (Keating, 2014a): strongly disagree (-3), disagree (-2), somewhat

disagree (-1), neutral (0), somewhat agree (1), agree (2), strongly agree (3). Each item or

statement represents one perception. The second section has something to do with the

second research question, and contains seven survey items about what the district did to

implement the CASE policy. Section three comprises four open-ended questions. The

first open-ended question pertains to the policy impact, which is the third research

question and will also be asked during qualitative case studies. The second one will

"give the respondents the opportunity to raise issues not covered by the attitude scales"

(Avramidis, Bayliss, & Burden, 2000, p.205) and to ask questions about districts'

implementation practices. The last part askes eight questions to collect demographic

information about district leaders and their districts.

I conducted a pretest to improve the survey instrument. This testing is supposed

to "reveal whether people understand and if they can answer the survey questions" (Fink,

2013, p.7), help "make the survey run smoothly" (Fink, 2013, p.8), and will be used to

"determine whether the questionnaire works in the manner intended by the researcher" (Czaja & Blair, 2005, p.20).  It is supposed to ensure language clarity, which is crucial to self-administered questionnaires (Fink, 2013), identify questions and problems like "unwarranted suppositions, awkward wordings, or missing response categories" (Presser et al., 2004, p.110), and ultimately enhance survey quality (DeMaio, Rothgeb, & Hess, 1998).  A conventional pre-testing method is the most relevant.  During a conventional pretesting, researchers "administer the questionnaire as they would during the survey proper" and after all responses are complete, researchers evaluate the problems reported by the participants (Presser et al., 2004, p.110).  As I distribute the survey online, I conducted the conventional pre-testing online.  I invited a small group of peers, campus and district leaders who have varying knowledge about the CASE policy, to review the survey instrument.  I carefully evaluated their feedback made some changes to improve the survey draft.

In addition, I conducted a survey content validity study because this exploratory study examines a brand-new policy and existing survey templates are not available to be modified for this study.  Although it is time-consuming, a content validity study is essential to producing accurate survey information and improving the validity of my study.  Rubio, Berg-Weger, Tebb, Lee, and Rauch (2003) outline procedures to perform a survey content validity study, which is the basis of conducing my content validity study.  According to Rubio et al. (2003), survey content validity study involves asking five to ten subject matter experts to check whether the survey asks appropriate questions to answer research questions.  Following the suggestions from Rubio et al. (2003), experts will be

asked about clarity of survey items and revision suggestions of measures. For my survey content validity study, I identified and contacted six subject matter experts in the fields of accountability study, policy implementation, and survey methods to participate in the content validity study. Half of them are from my institution and the other half are external experts. I also included the voice of a district leader who holds a doctoral degree, has overseen implementing the CASE policy in her district, and was one of the pilot study participants (she was excluded from taking the final survey). After subject matter experts returned the content validity study survey, I incorporated their feedback to modify the survey instrument.

Prior to distributing the survey, I conducted a reliability test, which is a final step to the survey development (Keating, 2014b). The reason is that "a reliable survey results in consistent information" (Fink, 2013, p.10). Either an internal consistency reliability test or a test-retest study are commonly used. The conventional internal consistency reliability that involves several items to measure "one attitude or characteristic" will not be tested for my study that is "interested in responses to each item" (Fink, 2009, p.42). Therefore, a test-retest study is the only option for improving the reliability of the survey portion of this study. The test-retest reliability test will measure whether respondents' answers to survey questions are consistent on more than one occasion (Fink, 2009). The reliability will be

> computed by administering a survey to the same group on two different occasions and then correlating the scores from one time to the next. A survey is considered reliable if the correlation between results is high; that is, people who have good (or poor) attitudes on the first occasion also have good (or poor) attitudes on the second occasion (Fink, 2009, p.42).

By convention, if the correlation coefficient is less than 0.80, I will revise the survey

items that generate different answers and then retest the newly defined items. A small

group of three district administrators from a Texas public school district were recruited to

participate. These participants had been involved in implementing the CASE policy.

After these people returned their two surveys' answers, I entered the data into Statistical

Package for the Social Sciences (SPSS) for Pearson correlation analysis and generated a

coefficient that was 0.72, which is close to 0.80. I made no change to the survey and

used this version as the final survey instrument. Since test-retest reliability participants

have more knowledge of the survey instrument than others, I excluded them from taking

the final survey.

**The survey sampling**. The survey study targeted district administrators from a

randomly sampled sub-set of traditional public school districts in Texas. The TEA

classifies public school districts into nine categories using a number of factors (TEA,

2015a). The factors include enrollment, growth in enrollment, economic status, and

proximity to urban areas (TEA, 2015a). Categories range from major urban to rural and

charter school districts constitute a separate group (TEA, 2015a). Charter school districts

were not included for this study, because they often have alternative accountability

schemes. The eight district types this study primarily focuses on are: major urban, major

suburban, other central city, other central city suburban, independent town, fast growing

non-metropolitan, stable non-metropolitan, rural (see Appendix B for detailed district

type definitions). I explored differences in the policy implementation by district types,

given that theories in chapter 2 indicate that varying district contexts and implementers might result in various practices and different perceptions of the same CASE policy.

I used a stratified sampling method to randomly select approximately 300 districts from a total of 998 districts in Texas and then recruited one administrator from each district who was involved in the implementation of the CASE policy to take an online survey. A typical random sampling method is inappropriate as districts vary across the state in many ways and the proportion of different types of districts in the resulting sample are unlikely to replicate the population proportion of statewide districts. If the sample is not representative, the survey results cannot speak to the population and the value of this survey will compromise. The TEA's existing eight classifications of public school district provide a ready-to-use subgroup or strata and imply a stratified random sampling method. Fraenkel and Wallen (2009) define stratified random sampling

> as a process in which certain subgroups, or strata, are selected for the sample in the same proportion as they exist in the population…The advantage of stratified random sampling is that it increases the likelihood of representativeness, especially if one's sample is not very large. It virtually ensures that key characteristics of individuals in the population are included in the same proportions in the sample (p.94).

In light of the above definition, stratified sampling should be used to make sure that a representative sample of participants are obtained for my study.

The plan was to recruit 300 districts to work with the statistics properly, because this is thirty percent of the population and can generate survey results that speak to the population (Keating, 2014a). Each of the eight types of districts was used as a stratum or subgroup. I calculated the number of sampled districts by each district type in

accordance with the proportion of each district type in the population (see Table 2). The

distribution of the actual sample size by district type approximated the proposed sample

except for rural districts. Since the total sample size reached the goal of one third of the

population, my survey results, to some extent, reflect the perceptions of statewide district

leaders.

**Table 2**
*Sample Size in Each School District Type*

| Texas School District 2014-15 | Population | Percent | Proposed Sample | Actual Sample |
|---|---|---|---|---|
| Independent Town | 70 | 7.0% | 21 | 22 |
| Major Suburban | 78 | 7.8% | 23 | 37 |
| Major Urban | 11 | 1.1% | 4 | 5 |
| Non-metro Fast Growing | 32 | 3.2% | 7 | 11 |
| Non-metro Stable | 181 | 18.1% | 55 | 57 |
| Other Central City | 41 | 4.1% | 13 | 22 |
| Other Central City Suburban | 163 | 16.3% | 50 | 63 |
| Rural | 422 | 42.3% | 127 | 84 |
| Total | 998 | 100.0% | 300 | 301 |
| **Sample size** | **30% ≈ 300** | **300 =100*3** | | |

Note: Major urban and other central city districts are combined into one group (urban) when analysis is done given their small sample sizes. Non-metropolitan fast growing districts are also collapsed into non-metropolitan stable districts when doing analysis.

**Data collection.** The survey data collection spanned between mid-November

2015 and late July of 2016, during which I sent three rounds of survey requests to district

leaders in about 1,000 public school districts. By mid- February of 2016, I completed the

first round of email requests where I asked if they were interested in participating and

then sent out the survey link in the following email. When carrying out the second round

toward non-responding districts, I followed one survey respondent's advice to include the

survey link into the first request email. During the first and second rounds of email

distributions, I also checked the returned surveys on a regular basis to see if the surveys were complete and whether or not the respondents who agreed to participate had returned surveys. When surveys were incomplete due to technical problems, I worked with the survey respondents to collect the data. When the participant did not take the survey, I sent them a reminder email. By the summer of 2016, I counted how many more respondents were needed for each district type and started the third round. By the end of July of 2016 when no more surveys were returned and the resulting sample approached the proposed plan, I stopped the survey data collection process. In total, I distributed approximately 2,500 emails, since I contacted more than one person from the same district, had to send follow up emails to remind those who had agreed to participate to complete the survey, and requested people to retake the survey when their surveys were incomplete.

While I sent out request emails, I created three inventories in excel to manage returned surveys. Following the chronological order in which a survey was returned, I assigned a number to each district and indicated its district type. I entered people's contact information to another two spread sheets, one for thank-you cards and the other for gift cards. The purpose of the ten-dollar gift cards was to increase the response rate. Following survey study literature (Jabbar, 2014), I used small financial incentives in order to encourage a high response rate. It seems this incentive motivated busy district administrators to participate in my survey as more than one-third of the survey respondents received gift cards. The opportunity to share their perceptions and vent related concerns might also account for the relatively high response rate.

As Table 2 shows, the resulting sample is 301.  After data entry, I spotted few non-responses to a couple of Likert questions and assumed the questions might have slipped the respondents' minds as they answered all of the other survey questions.  I requested their answers via email and all of them answered.  While the total sample reached the proposed goal, the number of rural district leaders (84) does not match the proposed number (127), despite my efforts.  Many rural areas are one-school districts (K-12), where the principal is also the Superintendent.  For some reason, rural districts were reluctant to participate in my study.

**Data entry.**  I entered quantitative survey data associated with perceptions of the CASE policy (Research Question #1) into SPSS.  I first imported district background data provided by TEA into SPSS.  The TEA data contains information as of the academic year of 2014-15, including district codes, district names, district type, region number, the number of Title I campuses, total enrollment, the number and percentage of economically disadvantaged students.  I then added the quantitative survey data from the ten Likert survey questions into SPSS.  As aforementioned, when I identified a few non-responses to a couple of questions, I followed up by email right away and collected the missing data.

I organized the qualitative survey data from the multiple-choice question of 'who was involved in determining the rating criteria' (part of the Research Question #2 on policy implementation) into an Excel file.  Survey respondents completing my survey selected from a list of eight-category of actors to answer the multiple-choice question of 'who were involved in determining the rating criteria.'  I entered data to eight columns

70

and each row represented one district. In total, 302 district leaders returned their surveys with answers to the involvement question.

I put the qualitative survey data from the open-ended questions of the CASE impact (Research Question #3) into an Excel file. I created this excel matric table to examine how district leaders reported the policy impact differently. I entered policy impact data in one column and created columns to include data on three other open-ended questions associated with people's concerns and challenges, their questions for legislators, and their reflections on the CASE policy. Those data are consistent with survey respondents' answers to the policy impact question and can be used to identify the impact that was implicitly stated. A total of 299 survey respondents answered the CASE impact question.

**Analysis of quantitative survey data.** Using SPSS Version 22, I performed descriptive and inferential statistical analyses for the quantitative survey data on district leaders' perceptions of the CASE policy (Research Question #1) and reported findings in Chapter 4. The descriptive analysis is the cross-tabulation by district type, which helped identify the general pattern by each survey item and the pattern by district types. The inferential statistical analysis is for the exploratory purpose and based on theory and literature. As the discussion of the sensemaking theory in chapter 2, people's perceptions of the CASE policy might differ depending their district contexts, such as district type and academic accountability ratings. Further, literature implies that people from districts with high poverty schools are likely to perceive the CASE policy somewhat differently. The RAND implementation study of NCLB showed that urban schools are more likely to

be identified for improvement (Le Floch et al., 2007).  Thus, I may expect survey

respondents from urban districts to perceive the CASE policy differently than others.

Therefore, the inferential statistical analysis includes perceptions by district type and

district academic accountability rating (Table 3).

**Table 3**
*Analysis of Survey Quantitative Data*

| Dependent Variables | Independent Variable | Analysis |
|---|---|---|
| | District Type: categorical | Cross-Tabulation |
| | 1. independent town | |
| | 2. major suburban | |
| **District Leaders'** | 3. urban | ANOVA |
| **Perceptions, continuous** | 4. non-metropolitan | |
| 1. Strongly disagree (-3) | 5. other central city suburban | |
| | 6. rural | |
| 2. Disagree | | |
| 3. Somewhat disagree | District Academic Accountability Ratings: | ANOVA |
| 4. Neutral (0) | categorical | |
| 5. Somewhat agree | 1. met standard | |
| 6. Agree | 2. improvement required | |
| 7. Strongly agree (3) | 3. not rated | |

Note: The same analysis is for each survey item.

As Table 3 shows, I generated results by performing cross-tabulation and analysis

of variance (ANOVA).  The dependent variable is district leaderships' perceptions of the

CASE policy.  I treated the Likert scale data as treated continuous, which is consistent

with many studies (Albirini, 2006; Avramidis, Bayliss, & Burden, 2000; B. Cavas, P.

Cavas, Karaoglan, & Kisla, 2009; Shade & Stewart, 2001).  The two independent

variables are categorical and contains more than three groups, rendering the rational for

ANOVA.  Since ten survey items constitute the perceptual question, I ran the descriptive

and inferential statistical analysis by each survey question.  I did not recognize any

outliers and tested the assumptions of independence and normality, which are not

violated (Gravetter & Wallnau, 2008).  The observations within the sample are independent, as each of them represents one district and is given the same chance of taking the survey.  Given the sample size of 300, the normality of the population is assumed.  The graphic test of histogram also shows a normal distribution.

**Data coding and analysis of qualitative survey data.**  I manually coded the qualitative survey data about the types of people who were involved in determining the CASE rating criteria (the second sub-question of Research Question #2 on CASE policy implementation) in Excel and analyzed them in SPSS.  After completing the data entering in Excel, I created an additional column and filled it with the number of groups each district engaged in determining the CASE rating criteria and imported the Excel data into SPSS.  I conducted content analysis and presented emergent points (Avramidis, Bayliss, & Burden, 2000) right after the results of district leaders' perceptions of CASE.  Based on descriptive results, I reported the pattern of districts engaging district staff and principals, teachers, parents, and community members, and discussed the districts with most and least involvement of stakeholders.  For each pattern, I explored possible reasons by connecting to data from open-ended questions.

Likewise, I manually coded the qualitative survey data on policy impact (Research Question #3) in Excel and analyzed then in SPSS.  After the data entering in Excel, I conducted content analysis in two stages.  I first generated a column to summarize the impact on each row using yes or no, not codable or not sure.  Then I created a column to further group the yes- and no-category into 'no-impact/not sure,' 'positive,' 'little impact,' 'some impact,' and 'large impact,' coding them numerically

73

from zero to four for basic descriptive analysis. Few survey respondents talked about the impact direction and magnitude. Some just reported positive impact and others talked about the magnitude of the impact using descriptors of 'little/limited/slight/minimal/not much/very small/very little/slightly positive,' 'some/moderate/mediocre/neutral/fair/mild/not huge/not big/not very significant,' and 'large/significant and positive/very positive'. I categorized those impacts that did not specify the magnitude but contained the wording of 'awareness or opportunity, good, help, useful, important, favorable, or promoting' as 'positive.' Among the 299 responses, 285 of them are codable. I imported the codable data into SPSS and reported findings as the last section in Chapter 4. First, I generated a bar chart to lay out how many district leaders identified any impact of the CASE evaluation, how many reported null effect, and how many was unsure of the impact. Second, I graphed the five groups of 'no impact/not sure,' 'positive impact,' 'little impact,' 'some impact,' and 'large impact' in a bar chart focusing on the nuances of the last four categories. I used participants' words to provide rationales on how they perceived the impact in their ways.

**Qualitative Case Studies**

I conducted qualitative case studies to explore how public school districts determined the CASE rating criteria, which is the second component of Research Question #3. The case studies and survey data collection occurred simultaneously (Teddie & Tashakkori, 2009). In this section, I first discuss the rationale for the qualitative approach. Next, I describe the selection of research sites and participants, and

then I address data collection and analysis procedures, followed by an explanation of data verifications strategies. Finally, I discuss my role in conducting a qualitative study.

**The qualitative approach.** A qualitative approach is appropriate as it allows an examination of the relationship between a policy and the context in which it is implemented (Hays & Singh, 2011). Since my study explores how district contexts shape administrators' perceptions and behaviors when they implemented a new assessment policy, the qualitative method is appropriate. The definition of qualitative research also verifies its appropriateness for this study. It is defined as "a means for exploring and understanding the meaning individuals or groups ascribe to a social or human problem" (Creswell, 2009, p.4). This method allows for understanding how districts perceived the new policy, how they determined the rating, and what they perceived to be the impacts of the policy. Given the limit of pre-determined questions of a survey, the qualitative method is particularly good for investigating the political dynamic when a group of people were trying to come to a consensus on determining the CASE rating criteria.

Specifically, I adopted the method of qualitative case studies. Merriam (1998) points out that when researchers are interested in insight, discovery, and interpretation rather than hypothesis testing, qualitative case studies should be chosen. This study is not interested in hypothesis testing, but it attempts to uncover how and why various school districts implemented the new policy in their ways. The theory of sensemaking and the political model also confirm the benefits of case studies. The former posits that people have preexisting practices and worldviews (Spillane, 1999) and that contexts determine people's construction of understanding (Coburn, 2001; Coburn & Stein, 2006). The latter

75

implies that power dynamics between diverse actors are associated with surrounding contexts (Malen, 2006). Aligned with the implications of the theories, qualitative case studies will reveal how different districts understood and implemented the new policy in their respective ways, and what power dynamics looked like in these districts. Honig (2006) even posits that a qualitative research design is an important source of knowledge for policy implementation study. Thus, qualitative multiple case studies are apropos choices to answer the implementation process component of Research Question #2.

This study utilizes multiple cases, because the purpose is to understand how the CAASE policy is implemented in different contexts. The goal of these case studies is not to generalize how an average school district responded to the CASE policy, but to examine how districts under different contexts implemented the new state policy (Holme, Carkum, & Rangel, 2013). In particular, this study fits special features that characterize qualitative case studies: "particularistic, descriptive, and heuristic" (Merriam, 1998, p.29). With an intention to model these characteristics, my study will focus on a few particular school districts that will be purposefully selected, generate thick descriptions about the implementation process and the outcomes of the CASE policy.

**Research sites and participants.** Given the significance of varying district contexts, I intended to recruit three to five districts representing different locales to illustrate how school districts in different contexts respond to the mandate from the state. I managed to gain access to four school districts through submitting research proposals: one suburban, one major suburban, one urban, and one major urban district. Due to practical reasons, these districts are from central Texas and close by. Ideally, I wanted to

76

include a minimum of one rural district because rural districts have the largest share of statewide school districts. Unfortunately, I was unable to recruit a rural school district. One rural district leader responded to my email request noting that they lacked staff to implement the CASE policy and were not in a position to participate in my interviews. District leaders in small districts are likely to take up multiple roles, and thus their tight work schedules might also have constrained their participation in my study.

**Table 4**
*Participants Per District*

| # | Interview | District Name | In Order of The Interview Time | Sex |
|---|-----------|---------------|-------------------------------|-----|
| 1 | 1 | Roadrunner ISD | Deputy Superintendent | M |
| 2 | 2 | | Director of STEM | F |
| 3 | 2 | | Director of Humanities | F |
| 4 | 3 | | Fine Arts Director | M |
| 5 | 4 | Warriors ISD | Chief of Leaning & Teaching | F |
| 6 | 5 | | Director of Research & Evaluation | F |
| 7 | 6 | | Director of Instructional Technology | M |
| 8 | 7 | Bulldog ISD | Executive Director of Campus Leadership | M |
| 9 | 8 | | District Lead Counselor | F |
| 10 | 9 | | Career & Technical Education Director | M |
| 11 | 10 | Cougar ISD | Director of Family & Community Engagement | F |
| 12 | 11 | | Assistant Superintendent in assessment | F |
| 13 | 12 | | Evaluation Analyst | F |

For each district, I interviewed three to four district leaders involved in implementing the CASE policy, such as associate/deputy superintendents, directors of curriculum or evaluation, and directors of subject areas (Table 4). I reviewed district websites to find potential interviewees, and found that in each district there were about three to five people in the district leadership role in charge of curriculum and accountability. I planned to contact these people directly after I received the district's research approval. However, in reality, the direct contact via email did not work well in

one district.  I began the interview with the district leader who approved my research and

obtained a list of potential interviewees who did participate in my interview.  Another

district recommended that I start by speaking with two top district leaders who were

deeply involved in the CASE implementation back in the spring of 2014.  I followed their

suggestion and requested one reference for the third interview from the first participant.

For the last two districts, the office that approved my research scheduled all the three

interviews in their districts, respectively, and I did not contact the interviewees directly.

While four people from the four different districts completed my survey, three of them

also completed my survey.  I asked the two who took the survey before the interview to

reflect on a few questions to which the answers are interesting to me.  Their answers

contributed to the better understanding of the perceptions of the policy.  In total, I

conducted twelve interviews with thirteen participants, including a group interview with

two directors from the same office in one district (Table 4).  Eight of the participants are

females and most of them (11 out 13, i.e. 85 percent) have worked in their districts for

several years and involved in implementing the CASE policy.  For the purpose of

confidentiality, I use pseudonyms for participating school districts and positions to refer

to district leaders.

Most interviewees have been in their current positions for several years and they

are quite familiar with the CASE policy and its implementation, except for two

participants arranged by one district.  Having been in the district for one to two years,

these two have much knowledge of the CASE impact, but knew little about the

implementation process.  After the interviews, I requested to speak with more people who

had assisted in implementing CASE. Unfortunately, the district responded that those individuals had left their district and moved to other companies or districts. In the meantime, I attempted to interview the one who completed my survey but with no good luck.

**Data collection and analyses.** I collected interview data between mid-January of 2016 and mi-July of 2016. For consistency across interviews, I created an interview protocol that is aligned with research questions (see Appendix C for the protocol). Interview questions primarily focused on the process of how each district determined the CASE rating criteria and the perceived impacts by each district. Before starting field work, I pilot-tested the interview guide by talking to a small group of three campus and district leaders. With feedback from the pilot testing, I modified the interview protocol by rewording some questions and aligning all the questions question with the purpose of the study. I distributed the protocol to participants prior to interviews so that they had an idea of what questions would be asked. I then conducted in-depth interviews in person at each of the case study sites. Each interview was semi-structured and lasted between thirty to sixty minutes. I audio-recorded (with consent) all but one of the interviews and had them transcribed. For the person who declined to be recorded, I took detailed notes during the interviews and wrote up my field notes immediately after the interview. Through emails, I collected follow-up data for more information. I also collected and solicited documents such as some districts' rating criteria and rubrics, district timelines, and their working documents. I attempted to use both interview data and documents to portray an accurate picture of each participating district's implementation of CASE.

I manually analyzed the transcripts of interviews and started the process after completing the interviews. Using rich interview data, I presented four district memos for which I introduced the districts and interviewees, described the process of determining rating criteria, and discussed the policy impact. Consistent with previous studies, I checked and rechecked the transcripts to make sure the analysis is consistent with interview data (Holme et al., 2013). Recall that the purpose of qualitative studies is to make sense of the process of determining CASE rating criteria, which is the second component of Research Question #2. I presented the findings in Chapter 5 and concluded the chapter with cross-case comparisons.

**Addressing trustworthiness and credibility.** To enhance the quality of the proposed qualitative research, I adopted a number of trustworthiness strategies, suggested in Hays and Singh (2011). I used these strategies in advance of my study, during the data collection and analysis, and after the study's completion. First, I kept a reflexive journal to document my self-reflections in the research process (Hays & Singh, 2011). Second, I triangulated multiple forms of evidence to better support and describe findings (Hays & Singh, 2011). As a method of increasing validity, triangulation refers to information collection from more than one individual and using several methods (Maxwell, 2012). Triangulation from several sources is a good way to minimize research bias and move in the direction of the researcher's choice (Hays & Singh, 2011), although this strategy will not "automatically increase validity" (Maxwell, 2012, p.112). I selected a triangulation of data sources involving several participants' voices. Three district leaders with different positions serve as a strategy of triangulation. I also triangulated interview data

with survey data and documents gathered from district websites.  Third, the method of

data analysis will also enhance data quality.  I analyzed my data throughout the data

collection process.  Since research questions, data sources, and methods might change as

the study proceeds (Hays & Singh, 2011), simultaneous data collection and analysis will

better capture the changes and will not overlook important findings.  The fourth strategy

is member checking.  During the interview, I posed probe questions for clarification and,

after the interview, I sent follow-up questions to participants for more information and

clarification (Hays & Singh, 2011).

       **My role as a researcher**.  I conducted this qualitative study to achieve scholarly

goals of understanding how various school districts implemented the CASE policy.  I

neither perceive myself as an expert in this policy implementation study nor as a naïve

learner.  Rather, I approach this study with curiosity and some knowledge from literature

and a pilot study.  Informed by Hays and Singh (2011), if collected data are contrary to

my experience or expectation, I planned on being honest and open with the data to

represent the voices of my participants through revisiting the text and asking follow-up

questions.  My study did not generate conflicting data and reports all information

gathered from surveys and interviews.

**Limitations of This Study**

       This study has limitations related to its quantitative and qualitative dimensions.

First, with regard to the surveys, there are limitations associated with self-reported survey

data.  Robert (1995) notes that self-reported data are subject to biases due to sampling,

and reporting problems, and missing data.  My survey data might be affected by reporting

81

issues.  It is reasonable to imagine that some respondents might fill in a survey in a way

to make his or her district look good to the researcher.  Second, survey data and interview

data covered the policy implementation over the past three years, therefore, findings of

this study cannot speak to an over-time trend.  Given the scarcity of research on this new

policy, however, this study will provide an important first look at whether disparities

exist regarding district leaders' interpretations and implementations of the CASE policy.

Third, the survey data analysis does not provide causal connections between district

leaders' perceptions of the CASE policy and district contexts.  The quantitative analysis

of various district leaders' attitudes can still provide a glimpse of how statewide district

leaders generally perceive the new policy.  Consistent with previous studies, I hope that

"this exploratory study will reveal patterns and provide variables" that might be useful

for future studies (Jabbar, 2014, p.42).  Finally, qualitative case studies only cover four

school districts located in central Texas that do not represent the eight district types

across the state.  This is determined by the nature of qualitative studies that do not aim to

generalize.  But findings from my qualitative case studies might reveal patterns that can

be used by future researchers to compare against districts from other geographical

locations.

**Strengths of This Study**

Despite the limitations cited above, this study has several strengths that are

important to highlight.  First, this study uses stratified random sampling and sampling

bias should not be a problem. The actual sample approximates the proposed one, largely

making the findings of this study reflect the population of district leaders in the state of

Texas.  This study thoroughly investigates the CASE policy, its impact, and the implementation.  The empirical evidence is significant and timely for Texas policy makers to assess the state's evaluation of new measures, for Texans to know of this new policy and for other states to learn from Texas in the post-NCLB ear.  Further, findings from this study will offer policy implications for states that are implementing or are considering an alternative accountability model with locally-defined indicators for their approach of ESSA.  The large-scale survey data this study will produce and its findings will be tremendously critical for states to make thorough considerations before they replicate the local accountability model adopted in Texas.  The qualitative case studies will allow for deep knowledge on how contextual factors shape diverse districts' evaluation of academic factors that cannot be quantified.  Finally, as previously mentioned, few studies have explored the new CASE policy of 2013.  We know less about how districts interpreted, accepted, and implemented the seemingly-positive accountability policy.  This study will take an exploratory first step into this terrain.

# CHAPTER 4 RESULTS OF SURVEY DATA

This chapter presents the results of the analysis of quantitative and qualitative survey data. The surveys were aimed at understanding how district leaders perceive the CASE policy (RQ #1), who was involved in determining the CASE rating criteria (RQ#2), and how district leaders perceived the impact of CASE (RQ #3). For the first question, I collected quantitative survey data by asking respondents to answer ten survey Likert-type questions from 'strongly disagree' to 'neutral' to 'strongly agree.' To explore 'who was involved in determining the rating criteria,' I asked survey respondents a multiple-choice survey question, which allows them to select from a list of eight groups of actors who had contributed to the rating criteria development in the academic year of 2013-14 (see Appendix A for a complete list of survey questions). The third question was open-ended for respondents to enter text in response to the policy impact question. As discussed in Chapter 3, data from the multiple-choice question and the open-ended question are qualitative by nature and the responses were analyzed with content analysis.

The sample size for quantitative and qualitative survey data slightly varies. 301 district leaders returned answers to Likert-scale questions, making the sample size for quantitative analysis matches the proposed sample, so that the geographic distribution of the surveyed districts surveyed closely represents the geography of the state (Table 2 in Chapter 3). The survey respondents come from six types of districts: independent town, major suburban, other central city suburban, urban, non-metropolitan, and rural districts (see Appendix B for the TEA definition of district types). The number of people who

answers the involvement question is 302, and there are 285 codable answers to the CASE impact question out of 299 responses.

This chapter presents the results in six stages. First, I report the overall perceptions of the CASE policy (Research Question #1) based on the survey quantitative data from the ten Likert survey questions. Second, I highlight the variations of perceptions of CASE by district types. Next I explore the differences of perceptions of CASE based on the inferential statistical analysis of the survey quantitative data. I then summarize the findings of district leaders' perceptions of the CASE policy. What follows are the findings from 302 respondents' answers to the qualitative survey question of 'who were involved in determining the rating criteria' (Research Question #2a). Finally, I present the results of the impact of CASE (Research Question #3) based on the survey qualitative data from open-ended questions.

**The General Pattern of Perceptions of CASE**

The first four survey questions sought to capture how district leaders perceived the significance of the "community and student engagement" (CASE) policy. Based on the descriptive analysis of survey data, most of the districts expressed support for the CASE policy (Figure 1). When asked to what extent they agree CASE is a necessary supplement to standardized testing, many district leaders agreed (72.1 percent somewhat agree, agree, or strongly agree), with 18.9 percent disagreeing (somewhat disagree, disagree, or strongly disagree). Similarly, the majority agreed CASE is a positive addition to the test-based accountability system (77.5 percent somewhat agree, agree, or strongly agree). Many reported that CASE should be factored into the accountability

system in the academic year of 2017-18 (72.1 percent somewhat agreeing, agreeing, or strongly agreeing).  It is noteworthy that when answering if CASE offers districts an opportunity to start a conversation among community members, the vast majority of the survey respondents agreed (with 90 percent somewhat agreeing, agreeing, or strongly agreeing) with just 5.7 percent disagreeing (the top right of Figure 1; see Appendix E for detailed crosstabulation of people's agreement or disagreement).

**Figure 1**
*Public School District Leaders' Perceptions of the CASE Significance (n=301)*
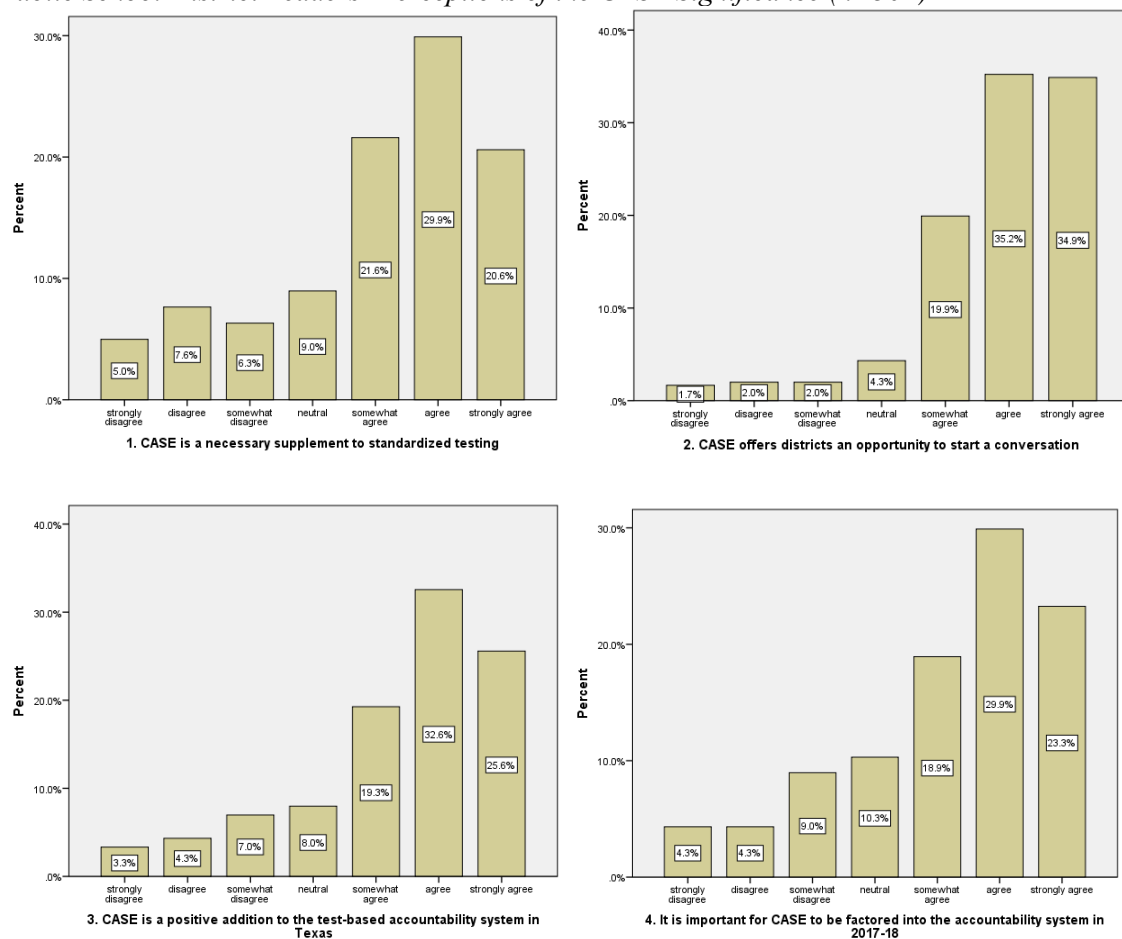
However, many district leaders felt they lacked state guidance or resources to implement the policy (Figure 2). When asked about their perceptions of the necessity of state guidance, most thought that the state should have provided local districts some guidance on rating criteria (75.4 percent of the surveyed district leaders somewhat agree, agree, or strongly agree). And most thought the state should have provided resources in support of local evaluation (75.3 percent; see Appendix E for detailed crosstabulation of people's agreement or disagreement).

**Figure 2**
*Public School District Leaders' Perceptions of State Support of the CASE Assessment (n=301)*
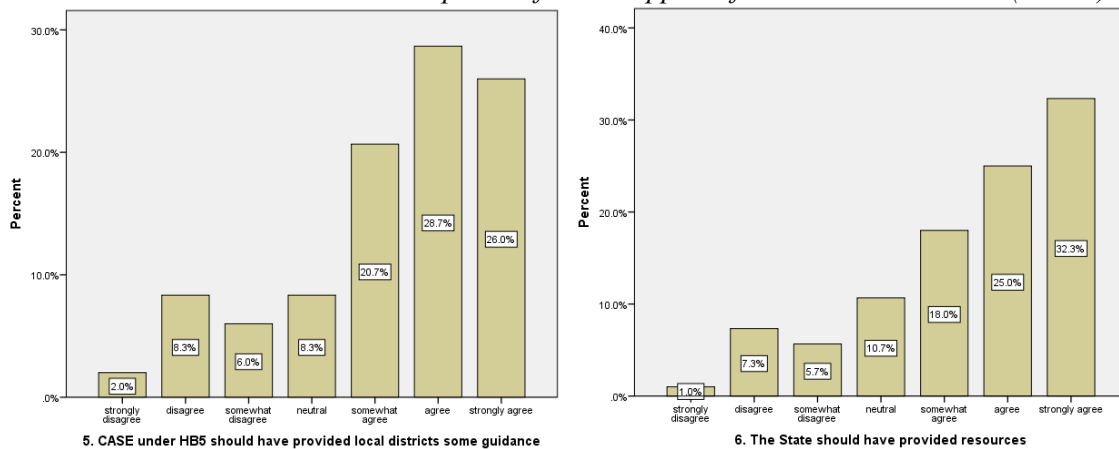


The last four survey questions were associated with the district leader's perceptions of the eight rating factors, state-assigned rating labels, and the process of developing rating criteria. Recall that under the CASE policy, districts must rate themselves on the required eight factors consistent with the state assigned labels of exemplary, recognized, acceptable, or unacceptable (see Appendix A for details). Most survey respondents agreed on the relevance of the rating areas and found comparing

ratings from self-evaluation to be problematic (the top of Figure 3). Many thought that all rating factors are relevant to their districts (with 86.7 percent somewhat agreeing, agreeing, or strongly agreeing). Since the rating criteria are locally defined, CASE ratings are incomparable across districts. Many people felt the lack of comparison is a policy weakness (with 76.5 percent agreeing and 15.6 percent disagreeing). Perhaps people have become used to standardization and being compared against one another under the existing accountability system.

There is much disagreement among survey respondents on the last two survey questions about the CASE rating labels and development process (the bottom of Figure 3). I asked all survey respondents whether they felt as though they should develop their own rating labels consistent with local rating criteria. I asked this question because, during the pilot study, some district leaders expressed they would like to develop rating labels in accordance with the locally developed rating criteria rather than using the rating labels required by the state. There is a lack of consensus about the need for local versus state labels. Slightly over half of the survey respondents (51.5 percent) thought so; almost 30 percent disagreed and another 20 percent held neutral perceptions. In the same vein, regarding the difficulty level of developing the rating criteria without any state guidance and one tight timeline, a little bit over half of the district leaders (62.4 percent) reported their districts found the process to be time-consuming and difficult. Given these difficulties, it is possible that the low-stakes CASE may not align with many districts' existing priorities or improvement plans and that some districts did not go to lengths to

reach a consensus on defining each rating factor and determining the rating criteria

assumed in the policy (see Appendix E for details).

**Figure 3**

*Public School District Leaders' Perceptions of The CASE Rating Factors, Labels, & Criteria (n=301)*
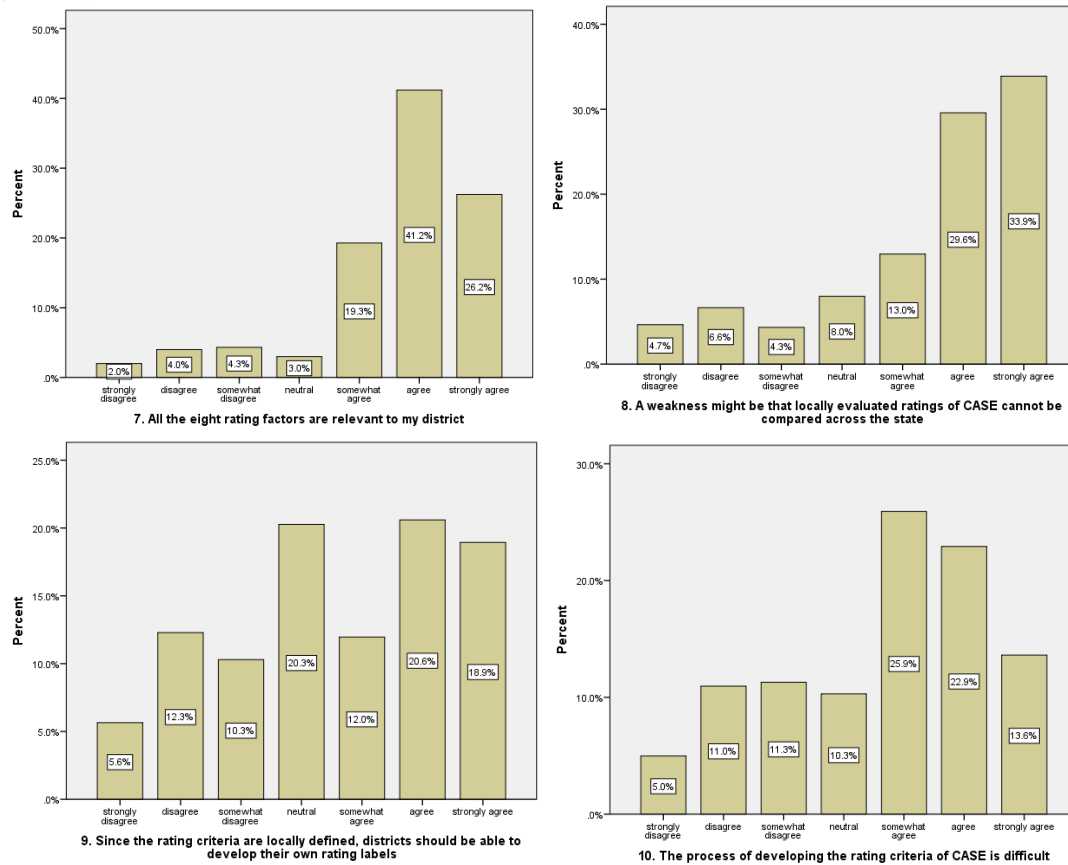


## Patterns of Perceptions of CASE by District Type

I next analyzed the data to see if there are differences by district type

(independent town, major suburban, other central city suburban, urban, non-metropolitan,

and rural.). I found there are distinct differences in urban, independent town, rural, and

major suburban districts compared to other districts on several survey questions (see Appendix E for details).

**Urban districts.** Education leaders from the urban districts (n=27) perceived the evaluation of CASE differently than their counterparts from all other districts on four survey questions (see Appendix E and F for details). First, urban district leaders were more likely to believe that it is a weakness that the ratings cannot be compared. It is striking that a large majority were dissatisfied that the CASE ratings across districts are incomparable. Almost 90 percent of the urban district leaders tend to agree (somewhat disagree, disagree, or strongly disagree) that the fact that CASE ratings cannot be compared across the state is a weakness while 74 to 77 percent of the other leaders from other types of districts see this way. This might have something to do with their reliance on and need for standardization to convince the public that their excellence on some areas of the CASE is reliable.

Most of them also wished that the state had provided support and found difficulty in developing their own rating criteria for the eight evaluation areas. For example, 88 percent of the urban district leaders believed that the state should have provided resources to help them assess the CASE, as opposed to 70 to 78 percent of other district leaders. They were more likely (70 percent) to report that the rating criteria developing process was difficult than other district leaders (about 60 percent). One possible reason is urban districts serving many low-income students lack resources and capacity to implement the CASE policy. This reason may also explain that urban district leaders were less inclined to find it necessary to discard the state-assigned rating labels and to develop their own

90

rating labels.  Only 44 percent agreed districts should develop their own rating labels that are aligned with local developed rating criteria, with 52 to 60 percent of other district leaders agreeing (somewhat agree, agree, or strongly agree).

**Independent town districts**.  The district leaders from independent town districts (n=22) were less likely to support the CASE policy as opposed to peers from other districts (recall these are districts that are non-urban, non-suburban and located in a county with a population of 25,000 to 99,999, with enrollment being the largest in the county or greater than 75 percent of the largest district enrollment in the county, see Appendix B for details).  Among the four survey questions regarding the significance of the CASE evaluation, independent town district leaders' agreement level is lower on three questions.  Around 46 percent of them believed (somewhat agree, agree, or strongly agree) that CASE is a necessary supplement to standardized testing scores.  This contrasts the percentage of district leaders from other types of districts ranging from 69.8 percent to 83.7 percent.  While 68 percent of the independent town district leaders thought that CASE is a positive addition to the test-based accountability system, almost 80 percent counterparts agreed (74.6 to 79.4 percent).  Likewise, 68 percent of the independent town district leaders reported that it is important to include CASE into the accountability system in 2017-18, with many more other district leaders agreeing (71.4 to 77.7 percent; see Appendix E and F for detailed patterns).  It is possible that people from independent town districts do not find the self-evaluation of CASE to be very meaningful or see the point of going that far.

**Rural districts**. I also found a different trend for rural district leaders (n=84) compared to the overall survey respondents on three survey questions. Like independent town district leaders, rural district leaders were less likely to support the CASE policy. While many (83 percent) noted CASE offers an opportunity to start internal conversations on how to improve schooling, the agreement level for leaders from other types of districts ranges from 90.5 to 94.5 percent. One possible reason is the rural districts are already doing what the law requires them to do, and they see the evaluation of CASE as less relevant.

Also, rural district leaders were less likely to want guidance from the state with regard to evaluating their performance on the CASE. While 67 percent of the rural leaders expressed (somewhat agree, agree, or strongly agree) that the state should have provided local districts some guidance, 77 to 85 percent of the other districts thought so. The fact that local control is more salient in rural schools may explain such disagreement.

Finally, like major suburban district leaders, rural district leaders did not find the eight rating areas to be relevant to their communities as much as their peers from other types of districts. 77 percent of the rural district leaders thought that all the eight rating factors required by the state are relevant, with 89 to 92 percent of other district leaders agreeing (see Appendix E and F for details). Given that business representatives weighed a lot in pushing the CASE to be incorporated into HB5 and there are fewer business in rural areas, it is possible that rural school districts are less likely to find the CASE assessment to reflect their concerns and values.

**Major suburban districts.** Finally, major suburban district leaders (n=37) demonstrate lower agreement levels than others on two survey questions. They are less likely to see the importance of factoring CASE into the accountability system in 2017-18. 68 percent of them tend to agree on the inclusion of CASE into the formal accountability as opposed to the 71.4 to 77.7 percent of district leaders from other types of districts. One possible reason is they did not find great value of the CASE self-evaluation that is less associated with student learning. Also, major suburban leaders along with their urban peers are less likely to agree on self-developing rating labels rather than using the state-assigned categories, with 30 percent agreeing (somewhat agree, agree, or strongly agree) as opposed to the 52 to 60 percent of other district leaders (see Appendix E and F for detailed patterns by district type). Suburban education leaders may find the consistency of state-required rating labels to be reasonable.

**Inferential Statistical Analysis of Perceptions of CASE**

Following the data analysis procedure detailed in Chapter 3, I examined the pattern of perceptions on the ten survey questions by district type and district accountability ratings for the exploratory purpose. The findings show the statistical differences only exist by district types and no significant results are found in the differences by district accountability ratings. One possible reason is that the self-evaluation of CASE is low-stakes and does not burden most districts.

**Table 5**

*Significant Results of Perceptions of Rating Factors and Labels*

5.1

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| **Corrected Model** | 34.317[a] | 5 | 6.863 | 3.651 | .003 |
| **Intercept** | 705.346 | 1 | 705.346 | 375.247 | .000 |
| **District Type** | **34.317** | **5** | **6.863** | **3.651** | **.003** |
| **Error** | 554.507 | 295 | 1.880 | | |
| **Total** | 1380.000 | 301 | | | |
| **Corrected Total** | 588.824 | 300 | | | |

Survey Question #7: all rating factors are relevant

Means by District

| District Type | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| Independent Town | 1.773 | .292 | 1.197 | 2.348 |
| Major Suburban | 2.081 | .225 | 1.637 | 2.525 |
| Urban | 1.926 | .264 | 1.407 | 2.445 |
| Non-Metro | 1.559 | .166 | 1.232 | 1.886 |
| Other Central City Suburban | 1.873 | .173 | 1.533 | 2.213 |
| **Rural** | **1.143** | **.150** | **.848** | **1.437** |

Survey Question #7: all rating factors are relevant

5.2

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| **Corrected Model** | 38.829[a] | 5 | 7.766 | 2.318 | .044 |
| **Intercept** | 57.058 | 1 | 57.058 | 17.029 | .000 |
| **District Type** | **38.829** | **5** | **7.766** | **2.318** | **.044** |
| **Error** | 988.427 | 295 | 3.351 | | |
| **Total** | 1129.000 | 301 | | | |
| **Corrected Total** | 1027.256 | 300 | | | |

Survey Question #9: districts should be able to develop their own rating labels

Means by District

| District Type | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| Independent Town | .727 | .390 | -.041 | 1.495 |
| **Major Suburban** | **-.216** | **.301** | **-.808** | **.376** |
| Urban | .222 | .352 | -.471 | .916 |
| Non-Metro | .882 | .222 | .445 | 1.319 |
| Other Central City Suburban | .508 | .231 | .054 | .962 |
| Rural | .821 | .200 | .428 | 1.214 |

Survey Question #9: districts should be able to develop their own rating labels

The differences by district type are statistically significant on two out of the ten

survey items (Table 5.1).  Significant variations are found regarding "all rating factors are

relevant" ($p = 0.003$).  Tukey post hoc results show rural district leaders are less likely to

find the rating factors to be relevant in their communities.  This is consistent with the

descriptive results in the previous section.  Specifically, the significant differences lie

between rural and major suburban/other central city suburban districts ($p = 0.008$ and $p =$

$0.019$, Appendix G).  As previously stated, the fact the eight factors under evaluation do

not speak to the rural community might have a lot to do with the CASE policy developing

process at Capitol where input from rural areas was not sought.

Significant variations are also found regarding "districts should be able to develop

their own rating labels" ($p = 0.044$, Table 5.2).  Tukey post hoc results also demonstrate

that fewer major suburban district leaders somewhat agree, agree, or strongly agree that

"districts should be able to develop their own rating labels." Specifically, the differences

between major suburban and non-metropolitan/rural are statistically different ($p = 0.041$

and $p = 0.049$, Appendix G).  It seems major suburban district leaders are more likely to

embrace the standardized state rating labels.

**Summary of Perceptions of CASE**

To sum up, descriptive results reveal much agreement on the first eight of the ten

survey items and a lot of disagreement on the last two survey questions.  Many district

leaders (more than 70 percent) agreed that the assessment of CASE is a necessary

supplement to standardized testing and a positive addition to the test-based accountability,

and that it is important to factor CASE into the accountability system in the academic

95

year of 2017-18.  Most survey respondents (over 75 percent) thought the state should have provided some guidance and resources to support the local evaluation of CASE and the lack of comparison of self-assigned CASE ratings is a policy weakness.  A large number of district leaders (86.7 percent) also felt that the eight state-assigned rating factors are relevant to their communities.  It is striking that the vast majority of them (90 percent) thought evaluating CASE offers an opportunity to start internal conversations among community members.  In contrast, 62.4 percent of the district leaders found it hard to regarding determine the rating criteria from scratch on tight timeline and about half of the survey respondents (51.5 percent) expressed they would like to develop rating labels in accordance with the locally developed rating criteria instead of using the state-assigned rating categories.

Regarding the variation of district leaders' perceptions of CASE by district type, descriptive and inferential analyses show interesting patterns for urban, independent town, rural, and major suburban districts as opposed to other districts.  Urban district leaders demonstrate differences in their perceptions of four survey questions about state support and CASE ratings.  Most agreed the state should have provided resources to help them assess the CASE (88 percent) and the process of determining rating criteria was difficult (70 percent).  While a great number of them (about 90 percent) were dissatisfied that the CASE ratings across districts are incomparable and saw it a policy weakness, they are less likely to agree that districts should develop their own rating labels that are aligned with local developed rating criteria (44 percent).  Independent town districts were less likely to support the CASE policy since their district leaders' agreement level on three of

the four survey questions regarding the significance of the CASE evaluation is low.

Fewer than half of them (46 percent) felt that CASE is a necessary supplement to

standardized testing scores. While 68 percent perceived that CASE is a positive addition

to the existing system and it is important to factor it into the accountability system in

2017-18, 70 to 80 percent of their peers thought so in these regards.

The variations of rural and major suburban district leaders' perceptions when

compared to other districts are salient in two to three areas and, some differences are

statistically significant. Rural districts are less likely to support the CASE policy, to want

state guidance, and to resonate with the eight rating factors. About 80 percent of the rural

districts agreed that CASE provides an opportunity for internal conversations beyond

high-stakes testing, but more than 92 percent of other districts thought so. Further, just

67 percent of the rural leaders noted that the state should have provided local districts

some guidance, while around 80 percent of the other districts agreed. Around 70 percent

of the rural district leaders found all the eight rating factors required by the state to be

relevant to their communities, with 89 to 92 percent of other district leaders agreeing.

Further, the differences between rural and major suburban districts, between rural and

other central city suburban districts are statistically significant. Finally, major suburban

district leaders were less likely to see it necessary for the state to factor CASE into the

accountability reporting system in 2017-18 (68 percent vs. 71.4 to 77.7 percent of other

district leaders) and for districts to self-developing rating labels. It is noteworthy that

one-third of the surveyed rural district leaders agreed districts should come up with rating

labels aligned with locally developed rating criteria, with more than half of the district

leaders in other districts agreeing (52 to 60 percent). The variations between major

suburban and non-metropolitan districts, between major suburban and rural districts, are

statistically significant.

**Who was Involved in Determining the CASE Rating Criteria**

Each district leader completing my survey selected from a list of eight-category of

actors (Table 6) to answer the question of 'who was involved in determining the rating

criteria.' The legislation requires districts to develop a local committee to determine the

CASE rating criteria. Based on the pilot study, people interpreted "local" differently and

involved various people in the criteria determining process. I created this list to include

possible implementers. As it may not be inclusive, I added an 'other' category for the

participants to provide more information. In presenting data, I divided the involvement

into internal and external resources to reflect nuances.

Since CASE evaluates the performance of campuses and districts on the eight

domains and one compliance factor, it comes no surprise that most districts saw central

office staff and principals as the primary groups to determine the CASE criteria. Among

the 302 surveyed districts, 95.4 percent of them (288 out of 302) had their central office

administrators involved (Table 6). Among the 302 surveyed districts, 90 percent of them

(272 out of 302) included their campus principals. Of the thirty districts that excluded

principals from the criteria developing process, half solely relied on central office staff or

an external model.

**Table 6**

*List of Actors Involving in Determining the CASE Rating Criteria (n = 302)*

| Who was getting involved | | |
|---|---|---|
| Internal Resources | | |
| 1. District staff/employees | 288 districts | 95.4% |
| 2. Principal Representatives | 272 districts | 90.1% |
| 3. Teacher representatives | 247 districts | 81.8% |
| 4. Parent representatives | 245 districts | 81.1% |
| 5. Business and community representatives | 219 districts | 72.5% |
| | | |
| External Resources | | |
| 6. Use/modify the toolkit by regional education service center (ESC) | 72 districts | 23.8% |
| 7. Use/modify the toolkit by Texas Association of School Administrators (TASA) | 84 districts | 27.8% |
| 8. Other: e.g. students or other districts or other professional organizations (TASB, etc.) | 28 districts | 9.3% |

Overall, I found that parents, teachers, and business and community representatives were less likely to be included than district staff and principal representatives (Table 6). I found that 81 percent of the total surveyed districts included parents in the rating criteria developing process. Some of the 19% of districts that did not include parents reported challenges such as the tight timeline, or outreach or transportation issues. One survey respondent from a district with 80 percent of economically disadvantaged students reflected on their challenges that "not all parents have transportation, baby sitters, clothes to wear, speak the language, value education." This seems to imply that some districts may not involve parents due to the districts' deficit view of parents.

Out of the surveyed districts, 247 (81.8 percent) of them included teachers, comparable to the proportion of parents. This means, however, that, 18 percent of the

districts (55 out of 302) did not include teachers into the process of determining the CASE rating criteria.  It is possible that those districts saw the non-academic measures under CASE to be less directly associated with teaching.  Thus, they decided not to burden the teachers who are already busy with teaching loads and serving other roles.  Some district leaders answering the open-ended questions about concerns also reported the challenge of engaging the "already over booked" teachers.

Finally, approximately 73 percent of the districts (219 out of 302) had business and community members participate in their process of determining the CASE criteria.  That is, more than 1/4th of the districts did not engage business and community representatives.  It is possible that those districts interpreted 'local' of the required local committee differently and had perceived community engagement to be more relevant to parents.  A deep look at the qualitative survey data shows that many of them are rural or non-metropolitan districts.  The absence of business and community representatives is possibly due to the fact there are few businesses.

**Figure 4**
*The Number of Groups Involved in Determining the CASE Rating Criteria (n = 302)*
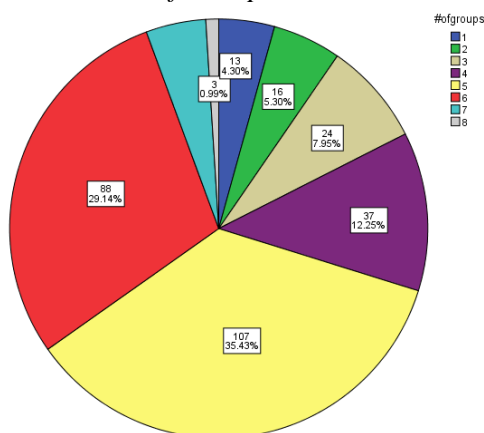
The analysis of survey qualitative data also shows variations of the number of groups each district involved in their process of developing the CASE rating criteria (Figure 4). Most districts involved four to six groups in the process of developing rating criteria, representing more than 77 percent of the surveyed districts. While slightly over 12 percent included three groups, more than 35 percent engaged five groups to develop the CASE criteria. It is possible that these districts had the capacities and resources to mobilize more people into the criteria determining process. There are thirteen districts that included seven groups (4.64 percent), and three districts that engaged eight groups (0.99 percent) into determining their CASE rating criteria. These districts seem to take CASE most seriously as they not only invited community members but also referenced to external models developed by professional organizations or listened to students' voices.

It is notable that a small number of districts engaged just one group or two groups of people in determining their CASE rating criteria (9.6 percent). Among the one-group involvement districts (13 out of 302, i.e. 4.3 percent), some just involved district staff, some only engaged principal representatives, and others rely on the external tool kit developed by the regional ESC or other professional organizations than TASA. For the two-group involvement districts (16 out of 302, i.e. 5.3 percent), most of them mobilized district staff and principal representatives and few of them included teachers, or used ESC or TASA as a resource.

**Perceptions of the CASE Impact**

I conclude this chapter with the results derived from the analysis of the qualitative survey data on the open-ended question of the CASE impact (Research Question #3).

Here I first present the general pattern and then examine the nuances of the perceived impact. I use data from other open-ended questions about people's concerns and reflections to provide possible reasons and details.

In general, most district leaders perceived that the CASE policy had exerted impacts on their districts (approximately 85 percent), few reported null effects (14 percent), and very few was not sure of the impact (Figure 5). For the 'no-impact' answers, some survey respondents explained their stakeholders did not take CASE very seriously, that it was just "another hoop to jump through for the state," and that the self-evaluation was "very subjective" and could not be compared with other districts. Others saw no impact because their "community and students are both already highly engaged in the school district" or their CASE replicates what they are doing. One survey respondent from a district where 91 percent of the students are economically disadvantaged noted that community involvement in his district is challenging and the implementation without resources from the state produced no change to their existing problem. Similarly, the aforementioned low income district with 80 percent of economically disadvantaged students saw no impact of CASE on their district as CASE is unfunded and they lack capacity to improve the engagement of the parents who have problems of transportation and babysitting. Another district leader saw the CASE implementation as an add-on and reported that principals did not really use CASE data and CASE was not an integrated part of district' or campus' work. Only six survey respondents were not sure of the CASE evaluation impact, and most of them just said they did not know. However, two people' explanations shed some light. One said, "I'm not sure if the public understands

102

the concept as of yet" and the other noted he or she was "not sure who reads or looks at it."

**Figure 5**
*The Perceived Impact of CASE (n = 285)*



The General Pattern of The Perceived Impact of The CASE Evaluation
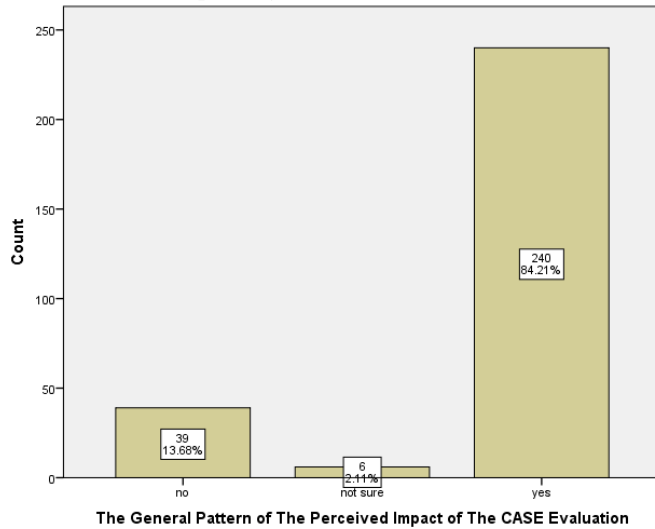
Further analysis of the impact category is displayed in Figure 6.  Approximately half of the sample (49 percent) reported positive impact without specifying the magnitude and more than one third (35 percent) talked about the impact in term of the magnitude, such as little impact, some impact, and large impact.  A couple of survey respondents reported both the direction and magnitude of the policy impact, such as 'slight positive' and 'little positive."  I chose to code them into the little-impact group.

In the positive-impact group, people explicitly or implicitly thought highly of the CASE evaluation impact.  Some explicitly stated the new policy was positive to their districts.  For instance, they noted CASE was the catalyst for them to craft a positive story about their district's performance to the community, a positive way to look at things they were doing besides just focusing on test scores, and "a positive experience for

103

campuses as they reviewed the data and criteria and staff members appreciate the opportunity to discuss programs and results that were beyond test results." Some districts saw it as "very positive" as their parents were heavily engaged in the affairs of the school district," a "positive move toward local control and the community members that are part of it like it," and a "positive tool for school staff and community." Others in the positive-impact group discussed the impact implicitly, using such descriptors as 'awareness,' 'opportunity,' 'good,' 'help,' 'useful,' 'important,' 'favorable,' or 'promoting.' Recall that in Chapter 3 I coded the responses containing these wordings into the positive-impact category. Specifically, some district leaders believed "it is a great opportunity for districts to reach out to existing students and community at large to hear what they have to say regarding our district" and that CASE "created an opportunity to have the staff and community come together to realize all the schools were already doing and look at ways … do more for students." Some noted that CASE "has created an awareness," "awareness of these activities at a campus level," and "it has benefited us by awareness of areas where we need to improve and it holds us accountable to make the changes."

**Figure 6**
*The Perceived Impact of CASE in Detail (n = 285)*
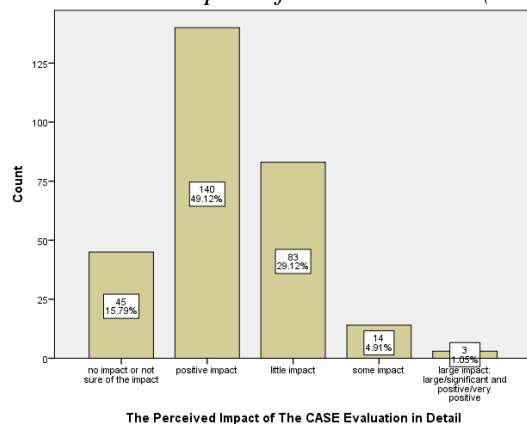
Supplementary data from other open-ended questions even shows that some people with positive views even expected the state to increase the weight of the CASE in the overall accountability system in the future, according to the survey qualitative data. For instance, one survey respondent noted, "I would encourage them to make this system an even greater component in state accountability ratings under future iterations of the law." Another district leader echoed, "Personally I would like to see the community and student engagement survey count for the lion's share of state accountability, with a smaller percentage of state test scores counting in the overall result."

Finally, I examine the impact-by-magnitude categories (little, small, and large impact) as they are illustrated in Figure 6. The little-impact category represents 29 percent of the surveyed district leaders whose answers were codable (83 out of 285) and contains some nuances. Recall that in Chapter 3 I coded all the responses describing impact as 'little,' 'limited,' 'slight,' 'minimal,' 'not much,' 'very small,' 'very little,' and 'slightly positive' into the 'little impact' category. Those who reported a very small impact offered different explanations: CASE is "just a compliance piece;" the self-evaluation is "too subjective" and cannot be compared across the state; "very few external entities take the self-evaluation seriously;" the districts "have always tried to be good in those areas;" they "were already concerned about the 8 domains prior to the implementation of the Local Accountability system" to name a few. Similarly, some people perceived minimal or little impact because they "were already implementing and evaluating" the CASE programs, CASE is "in line with (their) school improvement initiatives" and "just validates what are doing," their "community is already very

involved in the school," rating criteria are not standardized, and the public paid little

attention to the CASE evaluation. Another reason that the unfunded CASE assessment

did not "drive improvement" in the district. It is interesting to note that some reasons

listed here were interpreted differently by other district leaders who reported null impact

of the CASE evaluation (see the previous discussion of the no-impact).

The last two categories constitute a small portion of the survey respondents. The

small-impact group reflects the voices of less than five percent of the surveyed (14 out of

285, i.e. 4.9%) and the large-impact group reflects much less of the sample (3 out of 285,

i.e.1 percent). I find nuances in the small-impact category. Recall that in Chapter 3 I

grouped the data with descriptors of 'some,' 'moderate,' 'mediocre,' 'neutral,' 'ok,' 'fair,'

'mild,' 'not huge,' 'not big,' 'not very significant' into the 'some impact' category. One

district leader's reflection on the budget constraint on making changes based on the

CASE evaluation data may offer a reason why people reported some impact. The large-

impact group is the summary of three descriptors of 'very positive,' 'significant and

positive', and having 'large impact,' confirming the original purpose of CASE that allows

district to highlight their excellence and work for improvement. For instance, one shared

that they "used to have low participation from community members," which was changed

by evaluating CASE. Another two shared that CASE "is a very positive thing in that it

gives the District an opportunity to present all the truly great things our students and staff

are doing beyond the tests" and "the board of trustees and district employees appreciate

having a way to measure district performance in addition to the state mandated test. The

district staff works to make improvements for the goals set in each of the areas of this evaluation and progress has been made."

One caveat is that the impact reported here is a cross-sectional analysis, thereby making the positive impact not applicable beyond the data collection time. The CASE evaluation was implemented in the academic year of 2013-2014, right after its passage in the spring of 2013. The data collected occurred in the academic year of 2015-2016 when the CASE policy had been in place for less than three years. The same person might interpret the same fact differently and change their perceptions of the impact as time goes on. Survey respondents' then perceptions of the impact might change after 2017. This is because the 2017 85th legislature is considering an inclusion of a non-academic school climate domain into its accountability system in response to the ESSA requirement and CASE is one of the indicators that might be used. Regardless, the analysis of the survey qualitative data provides a glimpse of district leaders' perceptions of the immediate impact of the CASE evaluation at one point.

# CHAPTER 5 RESULTS OF POLICY IMPLEMENTATION

Drawing from the qualitative case study data, this chapter examines both the process of determining the CASE rating criteria and the impact of the policy. The CASE evaluation first started in 2013-14, right after its passage in the spring of 2013. By the time of data collection between fall of 2015 and summer of 2016, it was during districts' third year of implementing the policy.

To investigate the process of determining rating criteria and implementation, I conducted semi-structured interviews with four public school districts (see Table 7). Within each district, I interviewed three to four district leaders who were primarily probed to recount the process of developing their rating criteria. To protect the privacy of participating districts, I assign pseudonyms to the districts and use positions to refer to the participants in the following descriptions.

The CASE implementation has evolved over the three years by the time of the interviews. Given that the policy took into effect right after its passage in the spring of 2013, in the first year of implementation (2013-14), all districts worked against a tight timeline, defining the evaluated domains, determining rating criteria and rubrics, collecting, compiling, and submitting data to the TEA by August 8 of 2014. In the second and third year (2014-15 & 2015-16), local districts continued with data collection and submission and some even came up with improvement strategies. I describe below how the four participating districts (Table 7) determined the CASE criteria in the first year. I also explore if they made changes to better serving students and the community in

the following years, or if they just took it as an add-on and routinely collect and submit data to the TEA. I described the CASE implementation chronologically.

This chapter first presents four district case summaries derived from interview data. Each case summary starts with briefly introducing the districts and interviewees, describes the process of determining rating criteria, and ends with discussing the policy impact. This chapter ends with a brief cross-case comparison of the implementation of these four districts.

**Table 7**
*Four Participating Districts*

| District Name | Roadrunner ISD | Warriors ISD | Bulldog ISD | Cougar ISD |
|---|---|---|---|---|
| **District Type** | Suburban | Major suburban | Urban | Major urban |
| **Campuses** | 9 | 51 | 18 | 228 |
| **Students** | <10,000 | <50,000 | >10,000 | >50,000 |
| **Race/ethnicity** | White 70.7% Hispanic 12.8% Asian 11.1% | White 42.7% Hispanic 30.3% Asian 13.5% African American 8.6% | White 50.7% Hispanic 41.8% African American 3.6% | Hispanic 70.1% African American 22.9% White 4.8% |
| **Economically disadvantaged students** (state average: 59%) | <3% | >20% | >40% | >80% |

Sources: public information from districts' website and Texas Tribune.

## Roadrunner ISD

**District and interviewees.** Roadrunner ISD is a high-performing and small suburban district. As Table 7 shows, it serves over 7,000 students in nine elementary and secondary schools. It is a predominantly white district and has few low-income students,

with just 2.7 percent economically disadvantaged students.  The three interviews included

the Deputy Superintendent, two directors from the district's curriculum office (STEM

and Humanities areas), and the fine arts director.  Their years of working in the district

ranges from eight to twelve years as I spoke to them.

        **The process of developing the rating criteria** (See Figure 7).  As noted

previously, the law requires a local committee to be set up for implementing the CASE

policy.  Roadrunner ISD used its existing district leadership team (DLT) as the district

CASE committee, and this team includes more than 50 people.   There are certain district

leaders who must be on the committee: principals from the eight elementary and middle

schools, high school principals and assistant principals, two parents from each of the nine

campuses, two teachers for each campus, one parent professional, two business

representatives, and two other community members.  The Deputy Superintendent headed

the district CASE committee and he noted that he made the decision to use DLT to guide

the CASE evaluation as the team is already representative of the district because it

represents various stakeholders.

        The district CASE committee started their work of organizing a meeting with all

of the principals in the spring of 2014.  The goal of the principals meeting was to

overview the content of HB5 and discuss the evaluation requirements of the CASE.

During the principals meeting, the committee also went through their evaluation

guidelines that were developed and presented by the curriculum office.  The guideline

specified what categories or indicators would be evaluated in each of the eight curricular

areas.  The district did not make this format from scratch.  The curriculum office created

their own evaluation template based on a CASE toolkit developed by Texas Association of School Administrators (TASA) and finalized it over one meeting.

After the principals meeting, the district CASE committee created the eight subcommittees aligned with the areas under CASE, chaired by directors of specific subjects. Due to the district's practice, the district CASE committee did not dictate what subcommittees should do. It was at the chairs' discretion to select their members and organize meetings to "brainstorm" the rating criteria and rubrics. Subcommittees generally consisted of teachers, parents, students, business representatives, and community members to decide on rating criteria and an evaluation rubric. Each subcommittee had a minimum of six to ten people. The district has the tradition of valuing students' voice in decision making and a couple of students served on some subcommittees. The fine arts director, who is chair of the fine arts subcommittee, noted that his was "a pretty diverse committee" (the third interview, 8/31/2016). His generated a list of twenty to thirty candidates and those who were willing to work on tight timeline became the subcommittee members, including four to six art teachers from different grade levels and fine arts areas, two high school students "who were involved in multiple Fine Arts activities, and three parents "whose kids were or had been engaged in multiple activities" (email follow-up with the third interviewee).

The subcommittees' main task was to define criteria and evaluation rubrics for subcategories of each area. The interview with the fine arts subcommittee chair provides details about the process. Recall the fine arts evaluation is supposed to determine how well the district ensure sufficient fine arts facilities for students and engages the

111

community in fine arts activities.  The subcommittee chair broke down his group into small groups to "work on specific wording, requirements, standards, at their grade levels, or school levels" (the third interview, 8/31/2016).  The whole committee then came together to share what they have discussed. They spent much time working on the descriptors associated with fine art engagement and tried to "come up with some agreement on particular phrasing" (the third interview, 8/31/2016).  He allowed each person on the committee to voice their opinions and projected all the descriptors on the screen for everyone to read and provide feedback.

The fine arts subcommittee chair said that they went into heated but "communal" discussions regarding what descriptors to use to define the categories constituting the fine arts area.  As part of the fine arts category, they discussed the use of space as a measure of the quality of the fine arts offerings of the district.  There was some disagreement about the use of size to measure arts quality in elementary school.  Elementary principals did not like the requirement and removed the wording as they lacked spacious place on campus.  When the document came back to the Director, he put it back and the final document did have some wording to that effect.  He also gave an example about space for art classrooms.  Some members proposed specific numbers of per square footage, but he disagreed as space depends on art activities.  "Visual arts need more space than our music people need at that level.  It would be wrong to put in a certain square footage" (the third interview, 08/31/2016).  In the end, they used "adequate" to describe art space so that different art areas and different grade levels can use the general criteria to interpret their space differently and provide reasons.

After the subcommittees completed their tasks, they submitted their criteria and scoring rubrics to the district curriculum office. That office compiled rating criteria from all subcommittees. After the compiling process, subcommittee chairs presented their respective rating criteria and scoring rubrics for all areas to campus principals. Principals had more than five weeks' time to present rating criteria and rubrics to their campus leader teams (CLT) to seek input. Principals then shared with the curriculum office. Afterwards, the office spent approximately two weeks working with subcommittee chairs to finalize the criteria and rubrics across areas.

The curriculum office ensured that the criteria and rubrics across areas looked uniform by employing word limits in each descriptor. In some cases, the curriculum people ended up shortening the subcommittee's detailed version. The fine arts Director indicated that the version revised by the district lost some strong wording and he felt "quite disappointed about it" since his subcommittee's hard work seemed to be diluted (the third interview, 8/31/2016). He was given the opportunity to revise it, putting strong wording back to some places, and sending the updated version back to the district. In general, his team was "happy with the work in general, but not with the wording" (the third interview, 8/31/2016).

Finally, the district CASE committee reviewed and approved the updated version. After the curriculum office's and principals' review, the district CASE committee met with subcommittee chairs and approved the updated rating criteria. The Deputy Superintendent sent the document to principals and the district for them to collect data

and presented the rating criteria and rubrics for the CASE evaluation to the Board of Trustees for information purposes.

**Policy impact**.  While all the participants of the Roadrunner ISD saw the CASE evaluation was time-consuming, they reported that the process of determining the rating criteria and evaluating their performance on CASE generated ongoing and positive impact on the district and campuses.  They all felt that the CASE policy allowed the district to "look at ourselves in those areas, what we think our needs are, where do we fall short" (the third interview, 08/31/2016).  For instance, in response to the previous year's rating, the fine arts department initiated additional informative events beyond concerts in the second year.

Further, the district took advantage of the evaluation opportunity to set up goals of increasing outreach to ELL parents to improve this group's engagement in their children's schooling.  In the second year (2014-15), the district organized "a couple of district-level parents' nights for ELL population" (the second interview, 01/20/2016).  After two-year's implementation, principals and campus leadership teams received "enough valuable information" to improve on community engagement (the first interview, 01/19/2016).  In the third year (i.e., the time of the interview, 2015-16), "the campuses all want to pretty much hold their own" (the second interview, 01/20/2016).  In so doing, the campuses would score the exemplary rating in the community and parental involvement area.

*Figure 7*
**CASE Implementation Process of Roadrunner ISD**

**Local CASE committee**
District's existing leadership team
Led by Deputy Sup.
**5.** Deputy Sup. reviewed and approved the final criteria and informed Board; sent criteria to principals for them to collect data.

**Curriculum office:**
**1.** Drafted Rating format based on TASA's model
**4.** worked with subcommittees to finalize the criteria based on input from principals

**Subcommittees:** min. 6-10 members
**2.** Set their criteria & rubrics & presented them to principals

| Fine arts | Wellness & PE | Community & parental involvement | The 21st century workforce | 2nd language acquisition | The Digital learning environment | Dropout prevention | Programs for GT students |
|---|---|---|---|---|---|---|---|

**3.** Principals gathered input from campus leaders, shared with Curriculum office

## Warriors ISD

**District and interviewees.** Warriors ISD is a large and diverse suburban district. It has witnessed a lot of growth in terms of the number of students in the past five to ten years, changing from a moderate to a large district (the second interviewee, 2/18/2016). A possible reason is its location "in a very recognizably attractive place for business and families" (the third interviewee, 2/29/2016). Its student population is more than 45,000 students distributed among in their fifty-one campuses (Table 7). The district is ethnically and economically diverse as reflected in the more than 80 languages spoken by

their students and a mix of high- and low-SES families (the second interviewee,

2/18/2016). The majority of the students (42.7 percent) are white and Hispanics are the

second largest group (30.3 percent). The number of their economically disadvantaged

students is 26.8 percent, approximately half of the state average (Table 7). However, this

group has doubled over the past twelve years, when it was about ten to twelve percent

(the first interviewee, 2/15/2016). The three participants in my interviews were the chief

of learning and teaching at the time of the interview, director of research and evaluation,

and director of instructional technology. The first two have worked in the district for five

to twelve years by 2016. The third interviewee used to work in Warriors ISD, left it for

an ESC job, and went back in 2013. All of them noted that Warriors ISD already has

their academy system that parallels the graduation endorsements of HB5 and it was easier

for them align what they have with the high school endorsement requirement under HB5.

Partly due to the extra time, they spent two semesters on developing the CASE rating

criteria.

**The process of developing the rating criteria.** Warriors ISD created a new team

as their local committee for implementing the CASE policy, which was guided by the

chief of learning and teaching, and co-chaired by the director of research and evaluation

and the director of community relations. One co-chair (the director of research, who was

also my interviewee) is an expert in statistics and the other excels in parent and

community engagement. The chief of learning and teaching identified more than thirty

district staff for the district CASE committee. There were principals, assistant principals,

counselors, specialists in different areas, administrators from the two chairs' departments,

and experts in each of the eight evaluated domains under CASE (document solicited from the district).  While teachers and other community members were not included on the committee, this structure did not entirely exclude their voices, though.  The director of research noted that her district CASE committee based their initial work of defining community engagement and student engagement on survey feedback from teachers, parents, and community members and the parent surveys were administered in English, Spanish, Vietnamese, Chinese, and Korean, which reflected the diverse population in this district.

In the fall of 2013, the local CASE committee developed definitions of community engagement and student engagement and a list of guiding questions to facilitate this process (document solicited from the district).  They started with defining community engagement and student engagement and exploring what existing data they had, where to improve data collection, and where to start from scratch.  They also attended a CASE workshop organized by their local Education Service Center.  The chief of learning and teaching and the two co-chairs then brought in parents, teachers, students, business members and community members to a focus group to listen to and incorporate their understanding of community engagement and student engagement.  Finally, staff from the two co-chairs' departments worked together to brainstorm guiding questions and they presented their proposal to District Administration and Instruction Leadership (document solicited from the district).

The district CASE committee then utilized a range of external sources for training and references to prepare determine their rating criteria.  They invited Moak, Casey, and

Associates, a consulting firm that is expert in Texas school finance and accountability, to provide a training course on the content and state expectation of CASE. The CASE committee also considered input from the Texas Associate of School Board (TASB), Texas TPA, Texas School Public Relations Association, local ESC, and other neighboring districts (document solicited from the district). are professionals and one is a statistician. This district has a good relationship with the Moak and Casey organization and TXTPA and always look to the two for input. After reviewing others' models, the two co-chairs decided to develop their model that used use three indicators to measure campuses on each of the eight areas of CASE and primarily drew existing data for evaluation.

After the CASE committee's initial work on defining community engagement and student engagement, developing guiding questions, and determining the use of three indicators, the chief of learning and teaching organized eight subcommittees. Each subcommittee had eight to ten people and they included district staff, principal representatives, staff from the assessment department. Parents and other community member did not participate. The director of instructional technology, one of my participants, was chair of the digital learning subcommittee and his team included eight to nine people, including principal representatives, people from the unit of research and assessment, and other district staff. This subcommittee chair noted that while the chief of learning and teaching decided on who served on his subcommittee, he used his technology department staff as an extended committee and sought feedback from his internal people.

To establish goals for measuring each of the eight areas of CASE on three indicators, the subcommittees first worked independently and together with the local CASE committee to select and define five indicators (from which the district would then narrow down to three). This process took many meetings during a period of six weeks lasted until the end of the fall of 2013. Before generating possible indicators, they first reviewed video from the local ESC about the content of CASE and other districts' models and the local committee's proposed definitions of community engagement and student engagement and guiding questions (document solicited from the district). One co-chair (the director of community relations) asked the sub-committees to brainstorm ten indicators. The local CASE committee probed subcommittees to think through guiding questions, such as "does the idea measure our collective definition of engagement?", "Is the fair and will it allow every campus to do well?", and "will the idea be easily understood by our community" (document solicited from the district; the second interviewee, 2/18/2016). After each subcommittee developed their ten indicators, the district CASE committee brought in principals, asking to rank the five top most meaningful indicators for them. Asking themselves "what is the least burden on campuses that also returns the most valuable insights for the district?", the subcommittees then worked cut the list from ten indicators to five (the third interview, 2/29/2016). The chairs of the local CASE committee compiled all subcommittees' work and determined the scoring thresholds for each indicator by running different statistical models.

The last step was the narrowing of five indicators to three by a standing committee of the district. This occurred in the spring of 2014, when the district CASE

committee presented to and requested feedback from the district administration and instruction leadership team (document solicited from the district; the second interviewee, 2/18/2016). Then principals and other administrators gallery-walked the proposed five indicators of each area and provided feedback. A group of district level instructional administrators then convened to narrow the five to three measures for each of the eight categories. Later, they made an informational presentation of the CASE rating criteria for all areas to the leadership of Warriors ISD Council of PTAs. They also shared the work with the other groups associated with HB5 and applied the CASE rating criteria to the district's alternative campuses. Procedurally, the last step was that the local CASE committee informed the Board, which was not involved throughout the process.

The three-indicator framework created consistency across areas, but the process of deciding on five indicators was not easy for subcommittees. The chair of digital learning subcommittee noted that they spent "a lot of time" on deciding on which five indicators in their existing system that would collect meaningful and impactful data and they struggled to define what parents' engagement in digital learning meant (the third interview, 2/29/2016). When they presented at the local committee meeting, they received feedback from other subcommittees to improve their work. For some subcommittees like fine arts, the required three indicators did not encompass their area as they found it hard to use three or five areas to represent the many programs they wanted to highlight. The fine arts department had more than one hundred different programs and the community has been actively participating in almost all fine arts events. The fine arts subcommittee had "a tremendous number of indicators they wanted to include, but a lot

were unmanageable", such as "who is accountable for counting how many is participating" (the third interviewee, 2/29/2016). The fine arts subcommittee came with up twenty indicators, which were narrowed to five based principals' feedback and then to three indicators by district instructional administrators. Even though three were accountable on CASE, this subcommittee still used other seventeen indicators to collect data "in their own ways for their own reasons" (the third interviewee, 2/29/2016).

In addition to the aforementioned focus group, parents and other community members got involved in the second year. According the co-chair who is also the second interviewee, the district CASE committee invited a small number of community members to review the CASE rating criteria and listened to their suggestions for revision. These people included Title I school parents, parents who were regular volunteers, a couple of principals, and active business members. There were no significant changes, though.

**Policy impact**. All the participants agreed that the CASE had little influence on the district and they have not identified areas of improvement so far. As the co-chair (director of research and evaluation) noted, they had "put a lot of efforts but to reap little benefit" (the second interviewee, 2/18/2016). The chief of learning and teaching said that CASE allows them to tell people about their exceptional programs and show business and industry that they can provide well-rounded education. The subcommittee chair on digital learning noted that CASE helped the district create new and potentially useful data if the principals wanted to use it. Reflecting on her answers to my survey, the co-chair explained that after the first collaborating year when they defined success in each area and discussed the best ways to measure success, it now "feels like more of a disconnected,

compliance activity" … "an add-on rather than an integrated part of our work as a district"

(the second interviewee, 2/18/2016).

*Figure 8*
**CASE Implementation Process of Warriors ISD**



**Local CASE committee**
New Team, guided by Deputy Sup. &
Chief of Learning and Teaching

**Two Co-Chairs:**
1. developed definitions and guiding questions based on survey feedback from the community and conversations with community, presented to the district leadership team
2. referred to external sources, but developed their own process: 3 measures for each area
4. compiled subcommittees' work and made scoring methods, & presented to the district leadership team
7. presented to another district leadership team. Met with alternative campuses for accommodations. Informed the Board.

5. Gallery-walk & feedback from principals and other administrators

5. Gallery-walk & feedback from principals and other administrators

**Subcommittees: 8-10 members**
3. Decided on (with the local committee & principals) and presented their criteria on five measures to co-chairs
6. narrowed from five to three

| Fine arts | Wellness & PE | Community & parental involvement | The 21st century workforce | 2nd language acquisition | The Digital learning environment | Dropout prevention | Programs for GT students |

## Bulldog ISD

**District and interviewees.** Bulldog ISD is a small urban district and this district

serves more than 10,000 students in eighteen regular and alternative campuses (Table 7).

Slightly over half of the student population is white, but its growing Hispanic student

body has reached approximately 42 percent. The three interviewees were an executive

director of campus leadership, the district's lead counselor, and the career and technical

education (CTE) director. As discussed in chapter 3, the first two have worked in the

district for one to years during the time of the interview and therefore have a somewhat

limited view of implementation. The two shared a great deal of important information about the policy impact and a little about the implementation process. The CTE person, however, has been in the district for seven years at the time of the interview and he had extensive information about the process of determining the CASE rating criteria.

According to by the CTE director, Bulldog ISD experienced serious budget cuts around the fall of 2012 and the spring of 2013, before and when the CASE policy came into law. The district lost many teachers and "a huge number of central office staff." For instance, the curriculum unit was all gone except the third interviewee, who was working there, later assigned to the CTE office due to the loss of the CTE director. He accepted the task of leading the CASE rating criteria development and later was promoted to be the new CTE director. In his words, the district got "no crew at that time. Everyone had multiple jobs." He was no exception, juggling between curriculum and CTE related work.

**The process of developing the rating criteria** (See Figure 9). Like other districts, Bulldog ISD followed the statute and set up their local committee for the CASE implementation. The Assistant Superintendent decided to use an existing team, District Performance Council (DPC), as the district CASE committee to frame the work plan and the Assistant Superintendent served as the chair. The annually district-elected DPC consists of 46 people, including district staff and outside members. The DPC has about two thirds classroom teachers and one third non-teaching district and campus staff. The outside members include at least two business representatives, two parents of students enrolled in the district, and two community members; they participate in DPC through

123

nominations and self-nominations. The district determines this makeup of the DPC to represent their community diversity (district's online information).

The Assistant Superintendent, who was the CASE committee chair, was the key person pushing the CASE implementation. She assigned the CTE director to work with directors in departments of fine arts, dropout prevention, physical education, and other areas that parallel to the categories of CASE. In late summer to mid-fall of 2013, the director first worked on his own to develop a framework but found out "it was not easy. Very very difficult to sit there and come up with something in conjunction with everybody" (the third interviewee, 3/28/2016). Later the TASB released their model and the CTE director found the TASB format to be creative, containing a scoring mechanism and allowing people to expand. In late fall of 2013, the director worked directly with his supervisor, i.e. the Assistant Superintendent to rewrite and reformat the TASB model to make it relevant to their district.

Unlike the other districts described above, this district did not create any subcommittees to match the eight areas under CASE. The district CASE committee decided that the CTE director individually worked with each area director to come up with the criteria in the spring of 2014. While the reason is unknown, it is possible that the shortage of staff due to budget cuts might be the main reason. The CTE director worked with each director two times. In the first meeting, the CTE director explained what was needed and gave each area director a chance to review and prepare input. In the framework developed from the TASB model, the rating criteria for most areas were

complete. The CTE director left blank the areas which was beyond his purview does not excel. In his words, "a lot of areas, I am not the expert. So, I would rely on the expertise of the people that sitting on the table with you." During the second meeting, if the rating criteria was acceptable to area directors, their focus was if the content and scoring made sense. If the rating criteria needed to be filled, such as in fine arts, technology, dropout prevention, and the 21$^{st}$ century workforce, the CTE director relied on the experts and followed their ideas for definitions. He raised questions for clarification, but made no changes.

When the CTE director finalized the CASE rating criteria and scoring rubric, he handed the document to the Assistant Superintendent without seeking feedback from principals, teachers, or community members as he was not told to. The Assistant Superintendent presented the rating criteria to the board for information purposes and to campuses for principals to work with their site-based teams to collect data to self-evaluate on the required domains.

**Policy impact**. Implementing the CASE policy has generated a "positive" impact on the community and students (the second interviewee, 3/28/2016). According to the interviewee, after the CASE implementation, the community "started to have some really robust conversations about how we can engage our community to tell us what is important to them." The community strong interest led to the district's creation of the learner profile that customizes students' needs. The process of creating the learner profile involved three steps. First, the community developed a strategic plan in late spring of 2014, which is the first year of the CASE implementation. They

125

expected the district to redesign "the learning model to include customization and innovation at every level" so that the district becomes "home of the most inspired students" (district's online information).  The district board approved and adopted the strategic plan and created the learner profile based on data from their education summits, surveys, and forums involving a large number of teachers, students, business owners, civic leaders, and other community members.  Afterwards, the district would integrate actions and goals aligned with the profile into instruction.

*Figure 9*
***CASE Implementation Process of Bulldog ISD***
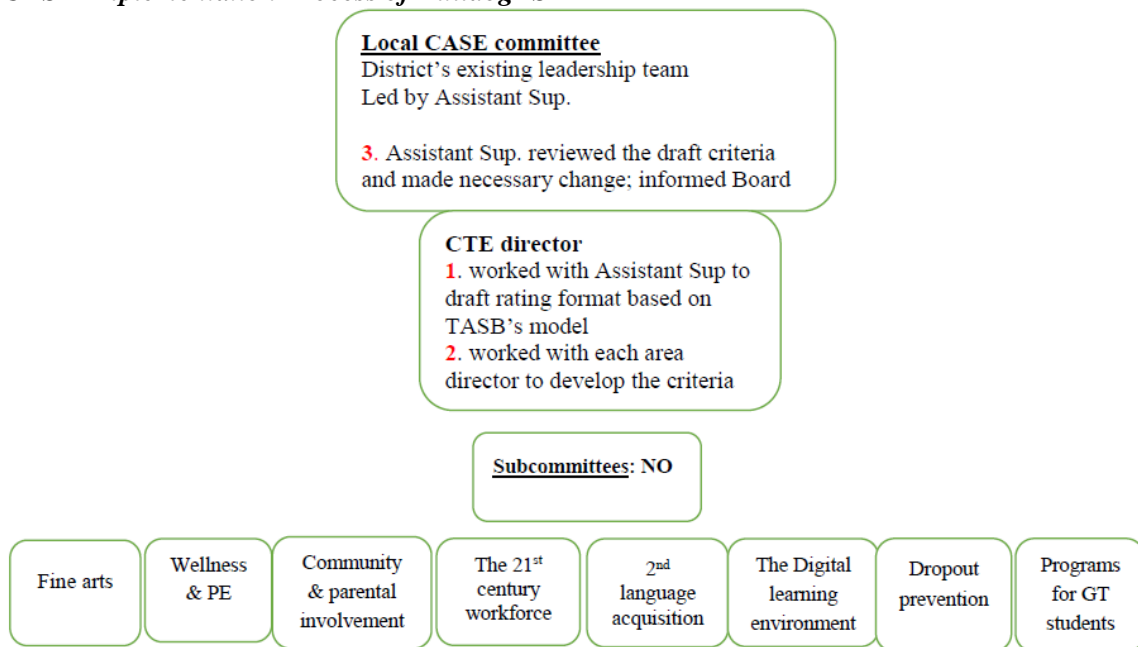


**Cougar ISD**

**District and interviewees.**  Cougar ISD is a major urban district, serving the largest student population among the four participating districts.  More than 150,000

students attend its over 200 campuses.  It is a predominantly Hispanic district with economically disadvantaged students being almost 90 percent.  The district's many initiatives at that time related to their new teacher evaluation system and the CASE policy was their low priority.  My interview participants are director of family and community engagement, the Assistant Superintendent in charge of the district's assessment office (short for Assistant Superintendent in the rest of this section), and her evaluation analyst. They have worked there for five to seven years by the time of my interview.

**The process of developing the rating criteria** (See Figure 10).  The Assistant Superintendent and her supervisor organized a new team as the local committee to implement the CASE policy, according to the evaluation analyst.  The two invited twenty to twenty-five representatives from all departments pertaining to the eight domains and other district staff, principal representative and other school leadership staff, and teachers. Leading the local CASE committee, the Assistant Superintendent started with assigning the evaluation analyst (the third interviewee) to review the content of CASE and come up with an idea on how to develop the rating criteria.  In the early spring of 2014, she organized a district planning meeting to present her plan to the central district leadership "so that they understand what they need to do" (the second interviewee, 7/18/2016). Recall that the state does not have any directions guiding local districts how to self-evaluate CASE.  Cougar ISD participated in their ESC region's work sessions, but the Assistant Superintendent and the Superintendent decided to start from scratch and "go with what the groups (departments) come up with" (the second interviewee, 7/18/2016). The goal was to "keep the principals' burden low" (the second interviewee, 7/18/2016),

engaging and supporting principals who primarily collect data for CASE (the first interviewee, 7/18/2016).

Like Bulldog ISD, Cougar ISD did not set up the subcommittee structure. The Assistant Superintendent charged the evaluation analyst to individually work with department leads responsible for a particular domain to define subcategories and determine rating rubric for each area. According to the evaluation analyst, "worked with each dedicated person assigned to each domain. If they did not get touch with her, she got in touch with them. And she … knows more about research, things that measurable. She helps them redefine, whatever their ideas, to make it workable." The department heads were asked to first establish their small working groups, comprised of department specialists and external partners if necessary, to collect their thoughts of the rating criteria for their domain (the second interviewee, 7/18/2016; district's document). Some departments engaged external organizations or listened to other districts (the second interviewee, 7/18/2016). The evaluation analyst then contacted the executive director of each department so that each division assigned a contact person to work with her. She drove around the district working with the contact person or a group of people sometimes to collect thoughts for the first version of rating criteria.

The eight departments' preparation work and passion about CASE varied a lot, leading to different collaborations with the evaluation analyst (the third interviewee, 7/18/2016). Some departments had many other things going on and could not get a person to sit down with the analyst, who had to call and worked with them on the phone

but was provided with two measures, "not very informed at all" (the third interviewee, 7/18/2016). In her words, "I have to finish the homework for them" (the third interviewee, 7/18/2016). Specifically, she turned to general district staff for brainstorming to develop the CASE rating criteria for a few areas that department heads did not contribute to. One department could not do it alone and accepted whatever suggestion the analyst offered. One department came up with new initiatives and saw their initiatives as a wish list. Although the evaluation analyst did not see these initiatives aligned with the district's goals, she modified their idea and added a measurable category. Some departments were very prepared, coming up with a long list of indicators to define their areas and designing a specific cut off for ratings. For instance, the fine arts department "basically designed their system" and made a presentation to the analyst (the third interviewee, 7/18/2016). The executive director of this department, the elementary and secondary directors, and other staff had the meeting together with the evaluation analyst to refine and tweak the rating criteria for fine arts. And the director of family and engagement (my first interviewee) brought her department's existing rating criteria of parent engagement to the evaluation analyst and both worked on the first draft. Parent and community engagement has been this department's priority and they already sought funding and sued their professional staff to develop their own criteria over the period of eighteen months before CASE was implemented. The director and the evaluation analyst combined the existing criteria with other pieces to make it reflect the CASE policy.

After meeting with all departments, the evaluation analyst drafted the first version

of the CASE rating criteria, sought feedback, and modified it to the second version in mid-spring of 2014.  In the first version, the evaluation analyst cut and combined detailed indicators for some areas to meet the district' and departments' goals.  Each area contained ten indicators and same scoring rubrics.  The evaluation analyst took the first draft to the local CASE committee and particularly listened to principals' feedback as they will be rated against the criteria.  Some principals engaged campus staff, teachers, and parents to provide their input.  The evaluation analyst worked with CASE committee members on the phone, rewrote the first version to make sure the rating criteria matched the domain title, and did "a lot of fix and revisions based on their suggestions," making the second draft of the CASE rating criteria (the third interviewee, 7/18/2016).

However, the CASE committee finalized its CASE rating criteria by not only basing on the second draft, but also referring to an external model into at the last minute. When the evaluation analyst presented the second version to the committee, the Assistant Superintendent, who led the local CASE committee, took the new draft to the district's principal focus group and districtwide advisory committee for input.  The former includes all principals and the latter contains parents, teachers, other community members. However, after Cougar ISD heard from other districts across the state that "give everybody high high marks" to make them look good in the eyes of the public, they were concerned that their rating criteria were "stringent" and "they will not do very well" (the second interviewee, 7/18/2016).  They decided to use another major urban district's data collection method of principal attestation survey that asked principals to confirm whether

or not their campuses met the criteria of the eight areas of CASE. The district asked their principals to complete surveys and tried out different ways of survey data analysis to ensure the scoring did not result in many exemplary ratings or unacceptable ratings. With the scoring method, Cougar ISD finalized their CASE rating criteria. Then the Assistant Superintendent informed the Superintendent of the final rating criteria, updated the Board, and provided the Board with results when available.

Cougar ISD basically did not change the criteria and scoring rubric in the next two years (2014-15 & 2015-16). Around December of each year, the evaluation analyst checks with each department to see if they want to do something different and workable. In the 2015-16 academic year, she updated the CASE rating criteria for parent and community engagement with family and engagement department's existing rubric that is a "higher standard" and part of the district's initiative, evaluating parents twice a year (first and second interviewees, 7/18/2016).

**Policy impact**. At the district level, the CASE policy did not produce any impact. This is because Cougar ISD did not use CASE data for improvement and unlike their teacher evaluation system, CASE is "not something really high on the radar" (the second interviewee, 7/18/2016). Regardless, the district continued with making the CASE evaluation system work better. In later May of 2016, which is two years after the policy implementation, the district changed the original paper survey for principal attestation to a digital survey (second and third interviewees, 7/18/2016). The electrical survey resulted in "way more less work" and principals "loved it" (the second interviewee, 7/18/2016).

The CASE evaluation also serves as a "a good summary" for the departments associated with the eight evaluated areas (the third interviewee, 7/18/2016). Department heads approached the evaluation analyst to check all the data and had a sense of "who (schools) isn't some of these things." "It is not news" to them, rather "it is useful to see each year they are improving" (the third interviewee, 7/18/2016).

*Figure 10*
**CASE Implementation Process of Cougar ISD**



**Local CASE committee**
A new team (not standing now) led by Assistant Sup.
2. The committee provided feedback to draft #1
4. Assistant Sup. took draft #2 to principal focus group and districtwide advisory group for input
5. Assistant Sup. combined another district' model to their draft #2 to make the final criteria & scoring rubric; informed the Sup. & Board

**Evaluation Analyst:**
1. worked with department heads of each domain and created criteria and scoring rubric draft #1
3. produced draft #2

**Subcommittees: NO**

| Fine arts | Wellness & PE | Community & parental involvement | The 21$^{st}$ century workforce | 2$^{nd}$ language acquisition | The Digital learning environment | Dropout prevention | Programs for GT students |

## Policy Implementation: Cross-Case Comparisons

To conclude, I briefly discuss the CASE implementation across the four participating districts. This analysis revealed three cross-case themes. First is the varied inclusion of stakeholders in these districts in the CASE implementation process. While the districts were implementing the same state mandate that requires a local committee to

define the CASE rating criteria, some districts took this meaningfully by engaging more stakeholders and others relied more on central office staff. Roadrunner ISD had extensively engaged the district and campus staff, parents, business and community members, and students into the CASE committee and subcommittees to determine the rating criteria. In contrast, Warriors ISD's district CASE committee was relatively more centrally organized, mainly engaging district staff, district specialists, and principals, they did engage the community in some ways. For example, their initial work was based on previous surveys completed by teachers, parents, and community members and parent surveys were administers in multiple languages to accommodate the diverse population. They also sought feedback initially from a small number of parents, teachers, students, and business and community and for possible revision of the CASE rating criteria in the second year from another small number of people that included Title I school parents, active parents and business members. Cougar ISD's local CASE committees and subcommittees were also more centrally organized, mainly including district staff and principals. However, Cougar ISD only sought feedback from a group of teachers, parents, and community members sitting on the district advisory committee to finalize he CASE criteria. Apparently, such engagement was very limited, and this district did not engage students, either. Finally, Bulldog ISD solely relied on a small group of district staff to determine the CASE rating criteria. It did not include voices of principals, teachers, parents, or community members.

Second is the differences of district contexts that were associated with the aforementioned varying inclusion of stakeholders. Roadrunner ISD is small and affluent,

133

has the tradition of including students and community members into decision-making, and their Superintendent emphasizes community's role in district affairs. It is not surprising that their inclusion of stakeholders was extensive and a range of actors participated in subcommittees. Warriors ISD's district leaders in implementing CASE are experts in statistics and community engagement and were familiar with their existing survey data from parents, teachers, and community members. Also, possibly due to their guiding principle of trying to reduce the burden on principals, they did not obtain much direct input from teachers, parents, and the community, but utilized professional district staff and principals to examine community feedback from past surveys and to select measures with existing data to define the CASE criteria.

While the low-stakes nature of CASE might be a common reason for the other two districts to handle the process centrally and to assign a district point person to work individually with an area director, their respective district conditions explain more of the variations. In Cougar ISD where their priority was on the teacher evaluation initiatives and there was low support of CASE, the district CASE committee only turned to teachers for feedback and did not engage parents, and community members throughout their implementation of CASE. The competing demands of teacher evaluation was the possible reason for this district to be compliance-oriented rather than engaging the whole community to internally examining their programs in the eight categories of CASE. Bulldog ISD's sole reliance on district staff and fewest inclusion of stakeholders had a lot to do with its budget cut before the CASE implementation, which resulted in a shortage of central staff and loss of teachers. Without sufficient human resources, it might be hard

for the district to reach out to the community, explaining what CASE was about and seeking community members' participation. As a result, this district based their rating criteria on an external model.

Third, the different policy impacts of the CASE implementation demonstrate another cross-case variation. Recall that the purpose of the CASE is to showcase areas of excellence and identify programs of improvement (see chapter 1). However, I found that only two of the participating districts have used CASE data or the policy opportunity for self-reflection. In Roadrunner ISD, the fine arts department initiated more events to engage parents and community members, both the district and campuses engaged in outreach to ELL parents by holding additional parents' nights. And in Bulldog ISD, during the second year of the CASE implementation, the community called for customized and innovative learning model and the district board engaged the whole community for input and created the learner profile to inspire their students.

The other two districts, however, seemed to view CASE as more of a compliance exercise, and made few changes based on the CASE implementation. Warriors ISD just reported ratings in the required programs without identifying areas for improvements, even though a participant noted that principals may find some CASE data useful for potential changes. Further, the interviewee who also took the survey reflected that CASE is not integrated into the district and she felt it "more of a disconnected, compliance activity" (second interview in Warriors ISD, 2/18/2016). Likewise, Cougar ISD just submitted CASE data to the TEA and did not utilize the data in a meaningful way except

135

that some of their district administrators checked CASE data to obtain a big picture of

campus performance in the eight assessed areas.

# CHAPTER 6 DISCUSSION

As discussed in Chapter 1, the CASE policy under HB5 of 2013 established eight non-academic factors and one compliance category for public school districts to self-evaluate their performance. The statute requires districts to set up a local committee to develop rating criteria, evaluate themselves, and then self-assign ratings to their district and each of their campuses of exemplary, recognized, acceptable, or unacceptable for both overall performance and each individual factor. They also must provide the state with an answer (yes or no) for district' and campus' compliance with statutory reporting requirements.

The CASE policy provides an instructive model of the movement in many states towards local accountability and more complex accountability systems that go beyond traditional measures such as test scores. Understanding this policy can also help contribute to the current national discussion of updating accountability in the wake of the ESSA. However, since its implementation in the 2013-14 academic year, few researchers have examined the CASE policy in terms of perceptions and implementation. This study was therefore designed to investigate how local districts perceive the CASE policy and its impact and how they implemented the policy.

As I noted in Chapter 2, after the passage of a policy, its implementation goes out of control of the intention of policy makers, and can play out in expected or unexpected ways. How a new policy is perceived and implemented largely depends on local contexts and implementers. There are multiple factors that can shape implementation. Recall from my theoretical framework in Chapter 2, sensemaking of new policies is associated with

one's social interactions (Weick, 1995) and contexts (Coburn & Stein, 2006; Spillane, 1996, 1998). In addition, according to the political model, policy implementation is a political process that reflects the power of diverse factors and the political dynamic is complex and takes different forms (Malen, 2006). Moreover, whether a policy is effectively implemented has a lot to do with the level of local institutional conditions and a balance of policy mandate and local support (McLaughlin, 1987). The new CASE policy is no exception; I found that local districts' perceptions of the new policy and its impact its implementation depended to some extent on each of these factors. Varying district contexts, politics, local conditions, balance or imbalance of mandate and local support mattered and play out differently in various districts and led to different policy impacts.

In this chapter, I first review the most prominent findings of district leaders' perceptions of CASE, the implementation pattern in four different districts, and the policy impact, as reported in the previous two chapters. I then discuss the theoretical and policy implications of this study, implications for practice, and suggest some future directions for research.

**Summary of Findings**

**Research question 1: How do public school leaders perceive the CASE policy?** To answer the first research question regarding district leaders' perceptions of the new policy, I analyzed quantitative survey data on ten questions. I found a wide range of support for the CASE policy overall. More than 70 percent of the district leaders surveyed came to a consensus that the assessment of CASE is a necessary supplement

and a positive addition to standardized testing, and that it is important to factor CASE into the accountability system in the academic year of 2017-18. 90 percent saw CASE an opportunity for them to start internal conversations and 87 percent perceived the eight rating domains to be relevant to their communities. However, over 75 percent of those surveyed thought that the state should have provided some guidance and resources, about 77 percent felt the lack of comparison of CASE ratings is a policy weakness, and more than 60 percent found the process of creating local criteria difficult. Finally, one half seemed to like the standardized rating labels for the local self-evaluation and the other half would like to develop local rating labels rather than using the state-assigned rating labels (exemplary, recognized, acceptable, or unacceptable).

Differences of perceptions by district type show more nuances and distinct patterns for urban, independent town, rural, and major suburban districts. Leaders from the urban districts tended to think that the ratings should be comparable across the state, rather than locally self-determined: almost 90 percent agreed that the fact that CASE ratings should be comparable across the state in contrast to 74 to 77 percent of the leaders from other types of districts. 88 percent of the urban district leaders believed that the state should have provided resources to help them assess the CASE, in contrast to 70 to 78 percent of other district leaders. They were more likely (70 percent) to report the rating criteria developing process was difficult than other district leaders (about 60 percent). They were also less inclined (44 percent) to discard the state-assigned rating labels and develop their local rating labels, with 50 to 60 percent of the other districts finding it necessary to come up with their own rating labels.

139

The district leaders from independent towns were less likely to support the CASE policy. Around 46 percent of them believed (somewhat agree, agree, or strongly agree) that CASE is a necessary supplement to standardized testing scores, which contrasts 70 to 84 percent of the district leaders from other types of districts. While 68 percent of the independent town district leaders thought that CASE is a positive addition to the test-based accountability system and it is important to include CASE into the accountability system in 2017-18, 75 to 80 percent of their counterpart agreed.

District leaders from rural areas, where local control seems to be already salient, were less likely to support the CASE policy, less likely to want guidance from the state, or less likely to find the eight rating areas under CASE to be relevant to their communities. Even though 83 percent noted CASE offers an opportunity to start internal conversations on how to improve schooling, 91 to 95 percent of their counterpart expressed this view. Only 67 percent of the rural leaders would like to have some guidance from the state while 77 to 85 percent of the other districts thought state guidance was necessary. And 77 percent of the rural district leaders thought all CASE domains to be relevant, with 89 to 92 percent of other district leaders agreeing.

Finally, major suburban district leaders were less likely to see the importance of factoring CASE into the accountability system in 2017-18 or agree on the need for discarding the state-assigned rating labels. Indeed, just 68 percent of the major suburban districts agreed on the inclusion of CASE into the formal accountability in contrast to 71 to 78 percent of the other district leaders. Also, only 30 percent of the major suburban leaders embraced the idea of locally developing CASE rating labels without using the

state-assigned categories, making them similar to their urban peers (44 percent), with 52 to 60 percent of other district leaders agreeing.

**Research question #2: How have public school districts implemented the CASE policy?** The second research question included two sub-questions. One focused on 'who was involved in determining the CASE rating criteria' and the other pertains to the process of determining the CASE rating criteria. The analysis of qualitative survey data indicated that more than 90 percent of the districts included central office staff and principals as the primary participants to determine the CASE criteria. This heavy reliance on district staff has something to do with the fact that CASE evaluates the performance of campuses and districts on the eight domains and one compliance factor. In contrast, parents, teachers, and business and community representatives were less engaged. Slightly over 80 percent of the surveyed districts included parents or teachers and around 70 percent of the districts engaged business and community members. The vast majority of the surveyed districts had engaged three to eight groups of stakeholders and external professionals into developing the CASE rating criteria (i.e. district staff, representatives of principals, teachers, and parents, business and community members, students, local ESC, TASA, other districts and professional organizations (see Table 6 in Chapter 4). A small number of districts engaged only one group or two groups of people in determining their CASE rating criteria (9.6 percent).

The analysis of interview data in the four case studies answers the second sub-question about the process for determining the rating criteria. First, the case studies show that a range of different stakeholders were involved. There were vast differences in who

were included into the process of determining the CASE rating criteria. Roadrunner ISD (suburban) included district staff, representatives from principals, teachers, parents, business and community members, and students throughout the implementation process. Warriors ISD (major suburban) primarily relied on district staff and principals and resorted to past survey feedback from parents, teachers, business and community members, directly engaging them once in the first year and once in the second year. Cougar ISD also mainly involved district staff and principals and only sought feedback from a small number of teachers, parents, and community members sitting on the district advisory committee in the CASE rating criteria finalizing phase. And Bulldog ISD engaged the least number of stakeholders, solely relying on district staff to determine the CASE rating criteria.

The case studies also illustrate varying policy implementation processes associated with different contexts. Roadrunner ISD, which was small, affluent, and has the tradition of involving community into decision-making, was able to draw wide participation into the process of developing the CASE rating criteria. Regardless, the district did not include all voices and one interviewee noted that while he invited a diverse group of people to his subcommittee, not all of them were willing to work against a tight time-line. In Warriors ISD, which tried to use as much existing data as possible for the CASE evaluation, specialists from the central office primarily worked with principals on developing the CASE rating with little engagement of teachers, parents, and community members. In Cougar ISD, various initiatives of teacher evaluation were the focus of district staff, principals and teachers and CASE became a low priority. This

district did not go to great lengths to engage teachers, parents, and community members to implement the CASE policy. However, to improve their rating criteria, this large district used a counterpart district's data collection method of principal attestation survey, which asked principals to confirm whether or not they met the criteria of each area of CASE. The budgets cut before CASE was implemented made caused a loss of many teachers and central staff in Bulldog ISD. The shortage of human resources mainly accounted for the district's decision to rely a small group of district staff to determining their CASE rating criteria without engaging principals, teachers, parents, and community members.

Finally, the case studies illustrate that there were vast differences in the impact of the CASE policy on district practices. I found that two of the four districts used CASE to identify areas of improvement, which is what was intended by legislators. With the district support, the fine arts department in Roadrunner ISD initiated more events to engage parents and community members. The ELL program director increased outreach to ELL parents by holding district-level parents' nights and motivating campus-level events. Similarly, in Bulldog ISD, despite the lack of community participation in the first-year implementation of CASE (2013-14), the community embraced CASE and asked the district to customize and innovate its learning model at different grade levels. The community's strong pressure pushed the district board to and create a learner profile in the second year (2014-15), which was based on community input and intended to produce "most inspired students."

In contrast, the other two districts viewed CASE more as a compliance exercise, as they didn't go beyond the CASE rating criteria development and data collection. They have not used CASE data to identify areas for improvement or go beyond the current practices in areas of community and student engagement. Interviewees from Cougar ISD and Warriors ISD noted that CASE is just a mandate and it is a "compliance activity," and neither of them engaged in initiatives in support of districts or campuses to make changes based on CASE data. Cougar ISD was occupied with their teacher evaluation initiatives as a top priority. And the Warriors ISD had not motivated campuses to integrate CASE data into their campuses' existing practices yet.

**Research question #3: How do public school district leaders perceive the impact of the CASE policy**? To answer the last research question of district leaders' perceptions of the CASE impact, I conducted content analysis of qualitative survey data as well as the qualitative case study data. The survey data revealed that there was a wide range in the impact of the policy. Sixteen percent reported null effect or they were unsure of the CASE impact. The majority of survey respondents identified some impacts (about 85 percent). Of these, approximately half of the surveyed reported positive impact, but did not specify the magnitude. Almost one third perceived the impact to be little or minimal, and a less than six percent noted some impact or large impact.

The survey data and qualitative case study data together showed district leaders' perceptions of CASE seem to shape the impact of CASE on their districts. Survey respondents who saw CASE only as a mandate, "another hoop to jump through for the state," tended to report null or very little impact. This is consistent with Warriors ISD

144

and Cougar ISD where leaders seemed to view CASE as a compliance exercise, and where they reported little influence on their districts or campuses and had not made any changes as a result of CASE. In comparison, many survey respondents who reported positive impacts tended to perceive CASE impact in a positive way, such as raising awareness and providing opportunities. Similarly, interviews at Roadrunner ISD and Bulldog ISD, had positive perceptions and reported positive impact of CASE on their districts.

Further, as noted in the qualitative case study data, the impact depends on district or community support. Roadrunner ISD supported improvement in CASE and they initiated changes that applied to all parents and ELL parents. There was strong community demand for making their district better in Bulldog ISD and the district board moved to make changes based on community inputs. In contrast, Cougar ISD prioritized many initiatives associated with teacher evaluation and district support for CASE is limited. As a result, this district reported little impact. Interviewees from Warriors ISD did not report the existence of district support and there was no change made by its district or campuses regarding CASE.

**Implications**

**Implications for theory**. As noted previously, I utilized several theoretical lenses to examine district leaders' perceptions of the CASE policy and their policy implementation process. Recall that I used sensemaking theory (Weick, 1995) to understand the perception of the CASE. According to this theory, contexts may affect how people perceive a new policy like CASE (Coburn & Stein, 2006; Spillane, 1996,

1998) and the analysis of the survey data reveals that context seemed to have an impact on how CASE was perceived.  For instance, most leaders from urban districts that are likely to lack capacity and resources to implement the self-defined and self-evaluated CASE policy agreed that state should have provide resource and that it was hard to develop the CASE rating criteria.  In the rural districts where there are few businesses and ELL students and much local control, it comes as no surprise that fewer district leaders found relevance in all the eight rating factors, some which pertain to the 21st century workforce development and the second language acquisition.

I used the political framework (Malen, 2006) to examine the process of CASE implementation.  As power dynamics among diverse actors are contingent on varying contexts (Malen, 2006), it is interesting to find out whose voices were included and who had the most influence on the way in which the process played out at different school districts.  The findings from both qualitative survey data and case studies demonstrate the varied inclusion of stakeholders, an indicator of different power dynamics.  Survey results showed a vast majority of districts involved district staff and campus principals but fewer engaged parents, teachers, and community members.  Interview data confirmed this, finding three of the four districts engaged district administrators and principals and that district leaders had more power of decision-making with regards to the CASE rating criteria.  Indeed, the inclusion of teachers, parents, and community members is only salient in Roadrunner ISD, which even included students' voices in many subcommittees. Warriors ISD included teachers, parents, students, and community members in limited ways.  The voices of teachers, parents, and community members were missing in Bulldog

146

ISD that lacked resources to mobilize the community to implement the CASE policy and in Cougar ISD that prioritized their teaching evaluation initiatives.

I also adopted McLaughlin's (1987) lessons from policy implementation to make sense of the process of determining the CASE rating criteria and policy impact of my case studies. Specifically, McLaughlin argues "internal institutional conditions" are critical for external policies to be implemented and that the balance or imbalance of policy mandate and institutional support accounts for different policy implementation patterns and results (1987, p.173). This framework relates to the limited involvement of stakeholders in Cougar ISD that prioritized teacher evaluation initiatives over CASE and in Bulldog ISD that experienced budgets cut and lacked staff to implement CASE. McLaughlin (1987) also argued the importance of a balance of pressure and support in effective implementation, which sheds light on the implementation results of the four participating districts. As previously discussed, Roadrunner ISD and Bulldog ISD reported positive perceptions of CASE and they made changes for improvement because of district and community support. These cases illustrate a balance of policy mandate and intuitional support. Warriors ISD merely perceived CASE as a mandate and showed little support because they were pressured by their teacher evaluation initiatives. Similarly, Warriors ISD saw CASE as no more than a compliance exercise and provided no motivation for campuses to generate changes from using CASE data. It is not surprising that such an imbalance of policy mandate and support produced negligible impact on these two districts and led to no changes for improvement.

147

**Policy implications.**  The examination of the CASE policy in my study provides timely information for Texas and other states who are makings revisions to their state accountability systems.  This study suggests that in order to effectively implement a new policy that is locally defined and evaluated, the state should clearly state its original definitions or guidelines that formulated the evaluated areas and provide some guidance and technical support.  The survey findings showed that many district leaders felt they needed state guidance as they found it hard to define the criteria and determine rating rubrics.  Many districts have competing demands and varying capacities and some districts struggled to define the eight areas of CASE.  A new mandate also means more extra work to the already limited time of their district staff.  Case studies suggested that some districts had to work overtime during after-school hours.  The self-defining process is challenging and time-consuming to almost all districts, including high-performing districts.  Some of my interviewees and survey respondents pointed out that districts have very diverse definitions of CASE, leading to high and low standards in different districts. Therefore, it is necessary for the state to act as a clearing house, directing districts to access external resources and toolkits developed by some professional organizations. The state may make some districts' best practices as examples for other district to improve their evaluation.

Findings from case studies show distinct patterns of determining ratings in different types of districts.  As such, the state can also provide more guidance about the implementation process.  This type of guidance should be tailored to types of districts and

made workable in various contexts. For instance, implementation guidance for urban districts that serve a large number of minority students should be differentiated from one-school rural districts where student population is small, superintendents are also principals, and staff undertake multiple responsibilities. Ultimately, the implementation can be compared across the same types of districts, rendering more rigor and value to the self-evaluation of CASE.

To ensure equity, the state should consider allocating funding to the districts that lacked capacities to implement the new policy or the districts that need resources to improve the status quo based on evaluation results. The survey findings also indicate that many district leaders felt they needed state resources to implement the policy, particularly the rating criteria development process. The accountability is problematic when some can develop better tools while other cannot. The accountability is meaningless when people identified areas for improvement but unable to make change due to the lack of resources, which is the dilemma of some districts in my study. Some district leaders from urban and rural areas reported that community involvement is a persisting problem, but it is beyond their budgetary means to change. As CASE is factored into the accountability system starting the 2017-18 academic year (HB 2804 of 2015; see Appendix A), holding districts with limited resources in policy implementation and policy improvement accountable becomes an imperative.

This study suggests that self-evaluation is subject to gaming in the same way as standardized measures. The state should think about the transparency of the self-

evaluation process and ensure the genuine self-reflection that lead to improvement.  One survey respondent shared that his district took CASE very seriously and was honest in the first year.  In the second year when a superintendent came in, they were questioned why they did not give themselves high ratings on all areas since it is a self-evaluation.  One interviewee of my case studies also noted that she learned many other districts rated themselves very high and she was pressured to tweak her district's scoring rubrics so that their ratings were not too low.

This study also offers implications for states to measure the non-academic school quality indicator required by ESSA.  Recall that the ESSA requires the assessment of school quality that can be one of several factors: student engagement, educator engagement, or school climate and safety (Klein, 2016).  The eight areas under CASE are relevant to the school quality factor and some states may find my study useful as they begin to design the non-academic ESSA factor.  In particular, Texas is considering using school climate for the non-academic school quality indicator (HB22 of 2017) and the CASE factors might play a significant role in the school climate domain if the state decides to utilize existing non-academic measures.  Filed by the chair of House public education committee, HB22 would charge TEA to develop new performance indicators under three domains: student attachment, school progress, and school climate.  Of the ten indicators that may be included to represent the school climate domain, one half would be teacher quality indicators, health and wellness indicators, and district or school climate evaluation determined by the Commissioner.  The fiscal note of this bill suggests that the

Commissioner may contract with a third-party to develop a statewide student survey to evaluate school climate.  The other half of the indicators are existing non-academic measures, including three programs of CASE that are selected by local districts from its eight areas and student course completion and participation records in certain courses and programs.  It indicates the three locally selected and evaluated CASE programs will take up at least half of the domain rating.  Recall that CASE is factored to into the accountability system in 2017-18.  It actually refers to the inclusion of three locally selected CASE areas (see Appendix A for details).  Unlike the Commissioner determined school climate evaluation that will cost extra funding from the state, CASE does not place any financial burden on the state.  It is in its fourth year of evaluation, has exerted positive impact on many districts, and would be the most optimal option for the school climate domain.

**Implications for practice.**  Interview data and district leaders' self-reflection in answering surveys show that few districts opted to give themselves high ratings on all evaluated areas.  As such, the state should ensure superintendents have a transparent process of implementation and make it accessible to the public.  Currently, many districts post their ratings on their websites., which means that the process of determining rating criteria and the content of the rating criteria are not available to the community or other districts.  If the implementation process and rating criteria are made public, districts may be more likely to take CASE more seriously and seek to improve relevant areas. In addition to enhancing transparency, CASE ratings would also be able to be qualitatively

compared across districts, which would motivate districts to continuously examine their performance at the district and campus levels.

The state should also work to ensure districts are genuinely inclusive of district and community stakeholders. When CASE passed the Texas legislature, researchers warned that varying capacities of each district might lead to problematical inclusion of local stakeholders in policy implementation (Lopez & Valenzuela, 2013). My study identified this problem and both survey data and case studies found the varied inclusion and exclusion of stakeholders in the process of determining the CASE rating criteria. To recap, district staff and principals were more engaged than teachers, parents, and community members. A small number of districts even either relied on district staff or adopted an external model to collect data without engaging any internal stakeholders. To make the local accountability really occur, the state may require districts to include stakeholders from diverse constituencies. Depending on the evaluated area (such as the second language acquisition program and educational programs for gifted and talented students) and the student population, immigrant parents, and/or parents of gifted and talented students, and/or low-income parents should be engaged into the implementation process. To achieve the end of including diverse stakeholders, it is critical that superintendents believe in and embrace community participation in implementing local accountability. Therefore, it is crucial for educational leadership preparation programs in colleges and universities prepare superintendents to adopt strategies for parent and community engagement.

152

**Future Research**

This study surveyed and interview district leaders, from assistant/deputy superintendents to directors in different areas, to understand the CASE policy. As I discussed in Chapter 3, it covers 3-year policy implementation without including all the proposed number of rural districts in the survey. The case studies on the policy implementation focused on urban, major urban, suburban, and major suburban districts due to access barriers. Therefore, while my study contributes to a better understanding of the new CASE policy and provide implications for policy and theory, this new policy warrants more research.

First, my survey respondents are top district leaders in one-third of the districts in the state. They represent their districts' general perceptions, but may not speak for their teachers, parents, or community members. Further studies may choose to survey district staff, teachers, parents, and community members to identify nuances of these people's perceptions. Second, future studies may explore CASE from the perspective of schools as schools are the primary unit of evaluation under CASE. Researchers may consider comparing and contrasting the CASE policy implementation and results in diverse types of schools, such as elementary and secondary schools, low SES or Title I versus high SES schools. Third, I call for research on longer-term implementation of CASE. As discussed in Chapter 3, this study covered CASE's three-year implementation and did not produce an over-time trend. As Texas is responding to the ESSA mandate, CASE might play a role in the assessment of school climate and districts' implementation of CASE might be updated. It will be intriguing to explore CASE from year one to the post-NCLB

era.  Fourth, this study does not cover rural or non-metropolitan districts.  Future researchers can explore these types of districts to see whether or not the theories used in this study are applicable and how the implementation patterns found in this study compare with that in other types of districts.

Additionally, the districts that most embrace the low-stakes CASE policy may be instructive to study.  Interview participants from one district noted that their neighboring district had monitored the CASE policy during the legislative session and started the implementation very early.  This type of districts are unique cases for qualitative studies as they may demonstrate very interesting implementation patterns and results from findings of this study.  Future research can also examine how district leaders' beliefs about community engagement affect their implementation.  Finally, researchers may experiment with other variables than district type and accountability ratings to uncover the variations of perceptions of CASE or similar local accountability polices.

# Appendix A: 'Community and Student Engagement'

In 2013, the 83rd Legislature of Texas passed House Bill 5 (HB5), a comprehensive bill that modified high school graduation plans and added the measure of 'community and student engagement' (CASE) and financial accountability to the state's existing test-based accountability system. The CASE policy applied beginning with the 2013-2014 school year. It is a key component of HB5 and currently in its fourth-year implementation. Its purpose is to allow each school campus to showcase areas of excellence in curriculum programs not reflected on standardized test scores and set improvement goals.

The statute requires nine factor that the district and each campus in the district must evaluate:
    (1) Fine arts,
    (2) Wellness and physical education,
    (3) Community and parental involvement,
    (4) Twenty-first-century workforce development programs,
    (5) second language acquisition programs,
    (6) Digital learning environment,
    (7) Dropout prevention strategies,
    (8) Educational programs for gifted and talented students,
    (9) Compliance with statutory reporting requirements.

This law gives no guidance on districts regarding which indicators should be used for each factor and the law does not permit Texas Education Agency (TEA) to determine the rating criteria. Rather, it requires districts to set up a local committee and engage the local community to determine the criteria for these factors and the district-wide self-evaluation.

Specifically, each district must use its own criteria developed by a local committee to rate the district's performance and the performance of each campus on the first eight domains (exemplary, recognized, acceptable, or unacceptable) and provide an answer (yes or no) for the compliance. The district and each campus should also assign a rating for overall performance and many districts post ratings to their websites.

Since 2014, districts must report ratings for the first eight domains and the compliance, as well as the overall ratings for campuses and the district, to the TEA. The deadline is August 8 of each year since 2014.

According to HB2804 of 2015, the evaluation of CASE is being incorporated into the accountability system. Starting the 2017-18 academic year, districts must select and report the evaluation and ratings for three of the eight programs to the TEA.

Here are the snapshots of ratings for a district and a campus in the district and the original texts of CASE in HB5 of 2013 and HB2804 of 2015.

**2015-16 Community and Student Engagement (CASE) Accountability Ratings**

**DALLAS INDEPENDENT SCHOOL DISTRICT**

| Domain | Rating |
| --- | --- |
| Fine Arts | Exemplary |
| Wellness and Physical Education | Exemplary |
| Community and Parental Involvement | Recognized |
| 21st Century Workforce Development | Recognized |
| Second Language Acquisition | Recognized |
| Digital Learning Environment | Exemplary |
| Dropout Prevention Strategies | Exemplary |
| Educational Programs for Gifted and Talented Students | Recognized |
| Compliance | Y |
| Overall Performance | Recognized |

Source:
http://www.dallasisd.org/cms/lib/TX01001475/Centricity/domain/98/data/hb5local/2016/2016_CASE_DALLAS_INDEPENDENT_SCHOOL_DISTRICT.pdf

**2015-16 Community and Student Engagement (CASE) Accountability Ratings**

**WOODROW WILSON H S**

| Domain | Rating |
|---|---|
| **Fine Arts** | Exemplary |
| **Wellness and Physical Education** | Exemplary |
| **Community and Parental Involvement** | Exemplary |
| **21st Century Workforce Development** | Exemplary |
| **Second Language Acquisition** | Recognized |
| **Digital Learning Environment** | Exemplary |
| **Dropout Prevention Strategies** | Exemplary |
| **Educational Programs for Gifted and Talented Students** | Recognized |
| **Compliance** | Y |
| **Overall Performance** | Exemplary |

Source:
http://www.dallasisd.org/cms/lib/TX01001475/Centricity/Domain/98/data/HB5Local/2016/2016_
CASE_WOODROW_WILSON_H_S.pdf

EDUCATION CODE
TITLE 2. PUBLIC EDUCATION
SUBTITLE H. PUBLIC SCHOOL SYSTEM ACCOUNTABILITY
CHAPTER 39. PUBLIC SCHOOL SYSTEM ACCOUNTABILITY
SUBCHAPTER C. ACCREDITATION

Sec. 39.0545. SCHOOL DISTRICT EVALUATION OF PERFORMANCE IN COMMUNITY AND STUDENT ENGAGEMENT; COMPLIANCE. (a) Each school district shall evaluate the district's performance and the performance of each campus in the district in community and student engagement and in compliance as provided by this section and assign the district and each campus a performance rating of exemplary, recognized, acceptable, or unacceptable for both overall performance and each individual evaluation factor listed under Subsection (b). Not later than August 8 of each year, the district shall report each performance rating to the agency and make the performance ratings publicly available as provided by commissioner rule.

(b) For purposes of assigning the performance ratings under Subsection (a), a school district must evaluate:

(1) the following programs or specific categories of performance at each campus:

(A) fine arts;

(B) wellness and physical education;

(C) community and parental involvement, such as:

(i) opportunities for parents to assist students in preparing for assessments under Section 39.023;

(ii) tutoring programs that support students taking assessments under Section 39.023; and

(iii) opportunities for students to participate in community service projects;

(D) the 21st Century Workforce Development program;

(E) the second language acquisition program;

(F) the digital learning environment;

(G) dropout prevention strategies; and

(H) educational programs for gifted and talented students; and

(2) the record of the district and each campus regarding compliance with statutory reporting and policy requirements.

(c) A school district shall use criteria developed by a local committee to evaluate:

(1) the performance of the district's campus programs and categories of performance under Subsection (b)(1); and

(2) the record of the district and each campus regarding compliance under Subsection (b)(2).

**Added by Acts 2013, 83rd Leg**., R.S., Ch. 211 (H.B. 5), Sec. 46(a), eff. June 10, 2013.

Sec. 39.0546.  PERFORMANCE IN COMMUNITY AND STUDENT ENGAGEMENT AS COMPONENT OF OVERALL DISTRICT AND CAMPUS RATING.  (a)  For purposes of including the local evaluation of districts and campuses under Section 39.053(c)(5) and assigning an overall rating under Section 39.054, before the beginning of each school year:

   (1)  each school district shall:

     (A)  select and report to the agency three programs or categories under Section 39.0545(b)(1), as added by Chapter 211 (H.B. 5), Acts of the 83rd Legislature, Regular Session, 2013, under which the district will evaluate district performance;

     (B)  submit to the agency the criteria the district will use to evaluate district performance and assign the district a performance rating; and

     (C)  make the information described by Paragraphs (A) and (B) available on the district's Internet website; and

   (2)  each campus shall:

     (A)  select and report to the agency three programs or categories under Section 39.0545(b)(1), as added by Chapter 211 (H.B. 5), Acts of the 83rd Legislature, Regular Session, 2013, under which the campus will evaluate campus performance;

     (B)  submit to the agency the criteria the campus will use to evaluate campus performance and assign the campus a performance rating; and

     (C)  make the information described by Paragraphs (A) and (B) available on the Internet website of the campus.

  (b)  Based on the evaluation under this section, each school district shall assign the district and each campus shall assign the campus a performance rating of A, B, C, D, or F, for both overall performance and for each program or category evaluated.  An overall or a program or category performance rating of A reflects exemplary performance. An overall or a program or category performance rating of B reflects recognized performance. An overall or a program or category performance rating of C reflects acceptable performance. An overall or a program or category performance rating of D or F reflects unacceptable performance.

  (c)  On or before the date determined by the commissioner by rule, each school district and campus shall report each performance rating to the agency for the purpose of including the rating in evaluating school district and campus performance and assigning an overall rating under Section 39.054.

**Added by Acts 2015, 84th Leg**., R.S., Ch. 1094 (H.B. 2804), Sec. 7, eff. June 19, 2015.


Source: http://www.statutes.legis.state.tx.us/Docs/ED/htm/ED.39.htm

# Appendix B: Survey Instrument

**Background**

As you know, House Bill 5 of 2013 added a non-testing measure –community and student engagement (CASE) –to Texas's educational accountability system. CASE establishes a rating system of nine factors to be locally defined and evaluated. This law also requires local districts to engage the local community in the rating criteria determining process.

**Purpose**

This survey is designed for district administrators involved in the evaluation of community and student engagement, as required by House Bill 5. I am interested in learning about your attitude toward the new non-testing evaluation system and your district's implementation practices. This survey is part of my dissertation that examines community and student engagement throughout Texas and in select school districts. I would greatly appreciate if you would complete this survey. It will take about 10-15 minutes to complete.

**Significance of this Survey**

I believe that the results of this study will benefit students, and educators, and can ultimately inform future policy decisions in Texas by providing insight into the specific aspects of districts administrators' attitudes regarding the evaluation of community and student engagement component of House Bill 5.

**Confidentiality Statement**

Your answers will be anonymous and will be reported only in aggregate form so that no individual's responses can be identified. Your individual responses will never be reported to anyone other than the researcher. All data will be collected and stored in ways that ensure individual responses will remain anonymous during the entire research process and thereafter.

**Compensation**

Your participation in the study is completely voluntary. Returning the survey implies that you understand your rights as a participant and voluntarily consent to participate in the study. You many ask any questions about the research at any time. A gift card for the amount of $10.00 will be awarded to you for your time and effort.

**Additional Information**

If you have questions about the research, please contact the Principal Investigator Dongmei Li by phone at (512)771-5228 or by email at dongmei_li@utexas.edu. If you are not satisfied with the response of the researcher, have more questions, or want to talk with someone about your rights as a research participant, please contact the UT Institutional Review Board at (512) 471-8871 or via email at orsc@uts.cc.utexas.edu.

**Section 1: 10 questions (Perceptions)**
Part A: The following 4 questions ask about your perceptions regarding the significance of "community and student engagement" evaluation.

| | strongly disagree (-3) | disagree (-2) | somewhat disagree (-1) | neutral (0) | somewhat agree (1) | agree (2) | strongly agree (3) |
|---|---|---|---|---|---|---|---|
| 1. Measuring "community and student engagement" is a necessary supplement to standardized testing scores. | | | | | | | |
| 2. The new policy offers districts an opportunity to start a conversation beyond test scores. | | | | | | | |
| 3. Overall, the new policy is a positive addition to the old test-based accountability system in Texas. | | | | | | | |
| 4. It is important for "community and student engagement" to be factored into the accountability system (in 2017-18 as required by HB 2804 of 2015). | | | | | | | |

Part B: The following 2 questions ask about your perceptions of resources in support of evaluation of "community and student engagement".

| | strongly disagree (-3) | disagree (-2) | somewhat disagree (-1) | neutral (0) | somewhat agree (1) | agree (2) | strongly agree (3) |
|---|---|---|---|---|---|---|---|
| 5. The law should have provided local districts some guidance on rating criteria. | | | | | | | |
| 6. The state should have provided resources to help districts evaluate "community and student engagement". | | | | | | | |

Part C: The following 4 questions ask about your perceptions of rating factors, ratings, and rating criteria development associated with "community and student engagement".

| | strongly disagree (-3) | disagree (-2) | somewhat disagree (-1) | neutral (0) | somewhat agree (1) | agree (2) | strongly agree (3) |
|---|---|---|---|---|---|---|---|
| 7. All the rating factors (i.e. see below) of "community and student engagement" are relevant to our districts and community.<br>(1) fine arts, (2) wellness and physical education, (3) community and parental involvement, (4) twenty-first-century workforce development programs, (5) second language acquisition programs, (6) digital learning environment, (7) dropout prevention strategies, (8) educational programs for gifted and talented students | | | | | | | |
| 8. A weakness might be that locally evaluated district and campus ratings cannot be compared across the state. | | | | | | | |
| 9. Since the rating criteria are locally defined, districts should be able to develop their own rating labels instead of having to use the state mandated labels. | | | | | | | |
| 10. The process of developing the rating criteria is difficult. | | | | | | | |

**Section 2: 7 questions (Implementation strategies)**

1. Who were included in determining the rating criteria of evaluating "community and student engagement" in the academic year of 2013-14? (You can select more than one)
❑ District staff/employees
❑ Principal Representatives
❑ Teacher representatives
❑ Parent representatives
❑ Business and community representatives
❑ use/modify the toolkit by regional education service center
❑ use/modify the toolkit by Texas Association of School Administrators
❑ Other (please specify) _____

2. What kind of data did your district collect to evaluate "community and student engagement"? (You can select more than one)
❑ PEIMS/existing data
❑ New survey data (please specify below)
❑ Other (please specify below)
    2.1 If you choose "new survey data" for question #2, what are the surveys? for which factor?
    2.2 If you choose "new survey data" for question #2, who developed the surveys?
    2.3 If you choose "new survey data" for question #2, who were surveyed?
    2.4 If you choose "other" for question #2, will you please explain?

3. Did your district post rating data to your district website?
◯ Yes
◯ No

4. Did your district examine the 2013-14 evaluation during the 2014-15 implementation?
◯ Yes
◯ No
    4.1 If you answer to question #4 is yes, did your district make any change?

5.  Did/Will your district improve campuses that were rated unacceptable?
◯ Yes
◯ No
◯ N/A
    5.1 If your answer to question #5 is "yes", how was (will) it (be) done?
    5.2 If your answer to question #5 is "no", can you explain why?

6. Will your district use the existing rating criteria for the 2015-16 implementation?
◯ Yes
◯ No
    6.1 If your answer to question #6 is "no", what changes does your district make?

7.  Does your district set goals where you want to grow?
◯ Yes
◯ No
    7.1 If you answer to question #7 is "yes", how are (will) your goals (be) achieved?

**Section 3. 4 Open-ended Questions**
#1. How do you perceive the impact of the "community and student engagement" evaluation?
 #2. What are your district's concerns/challenges, if any? (time input, associated cost, etc.)
#3. If you could talk to legislators, what would you ask them to change about the "community and student Engagement" evaluation?
#4. Is there anything else you want to share with me? ("community and student engagement" under HB 5 of 2013, HB 2804 of 2015, its evaluation, etc.)?

**Section 4: 8 Demographic questions.**
1. Please provide the FULL name of your district:
2. What is your gender?
❍ Female
❍ Male
❍ Prefer not to answer

3. What is your race/ethnicity? (You can mark one or more)
❍ White, non-Hispanic
❍ White, Hispanic
❍ Non-white, Hispanic
❍ Asian American
❍ Black or African American
❍ Mixed Race
❍ Other-please specify _____

4. What is your current job title/position at the district?
5. What is your role in evaluating "community and student engagement"?

6. You have been working in the public education field for
❍ 1-4 years
❍ 5-9 years
❍ 10-14 years
❍ 15-20 years
❍ 20+ years

7. Highest degree earned
❍ Ed. D/Ph. D
❍ ED. S
❍ M.ED
❍ M.S.
❍ Other (please specify) _____

8. In what year did you start your career in the public education field? PLEASE ENTER A 4 DIGIT NUMBER.

# Appendix C: TEA School District Type Definitions

"**Major Urban**. A district is classified as major urban if: (a) it is located in a county with a population of at least 870,000; (b) its enrollment is the largest in the county or at least 75 percent of the largest district enrollment in the county; and (c) at least 35 percent of enrolled students are economically disadvantaged. A student is reported as economically disadvantaged if he or she is eligible for free or reduced-price meals under the National School Lunch and Child Nutrition Program (TEA, 2015b)."

"**Major Suburban**. A district is classified as major suburban if: (a) it does not meet the criteria for classification as major urban; (b) it is contiguous to a major urban district; and (c) its enrollment is at least 3 percent that of the contiguous major urban district or at least 4,500 students. A district also is classified as major suburban if: (a) it does not meet the criteria for classification as major urban; (b) it is not contiguous to a major urban district; (c) it is located in the same county as a major urban district; and (d) its enrollment is at least 15 percent that of the nearest major urban district in the county or at least 4,500 students (TEA, 2015b)."

"**Other Central City**. A district is classified as other central city if: (a) it does not meet the criteria for classification in either of the previous subcategories; (b) it is not contiguous to a major urban district; (c) it is located in a county with a population of between 100,000 and 869,999; and (d) its enrollment is the largest in the county or at least 75 percent of the largest district enrollment in the county (TEA, 2015b)."

"**Other Central City Suburban**. A district is classified as other central city suburban if: (a) it does not meet the criteria for classification in any of the previous subcategories; (b) it is located in a county with a population of between 100,000 and 869,999; and (c) its enrollment is at least 15 percent of the largest district enrollment in the county. A district also is other central city suburban if: (a) it does not meet the criteria for classification in any of the previous subcategories; (b) it is contiguous to another central city district; (c) its enrollment is greater than 3 percent that of the contiguous other central city district; and (d) its enrollment exceeds the median district enrollment of 836 students for the state (TEA, 2015b)."

"**Independent Town**. A district is classified as independent town if: (a) it does not meet the criteria for classification in any of the previous subcategories; (b) it is located in a county with a population of 25,000 to 99,999; and (c) its enrollment is the largest in the county or greater than 75 percent of the largest district enrollment in the county (TEA, 2015b)."

"**Non-Metropolitan: Fast Growing**. A district is classified as non-metropolitan: fast growing if: (a) it does not meet the criteria for classification in any of the previous

subcategories; (b) it has an enrollment of at least 300 students; and (c) its enrollment has increased by at least 20 percent over the past five years (TEA, 2015b).”

**“Non-Metropolitan: Stable**.  A district is classified as non-metropolitan: stable if: (a) it does not meet the criteria for classification in any of the previous subcategories; and (b) its enrollment exceeds the median district enrollment for the state (TEA, 2015b).”

**“Rural**.  A district is classified as rural if it does not meet the criteria for classification in any of the previous subcategories.  A rural district has either: (a) an enrollment of between 300 and the median district enrollment for the state and an enrollment growth rate over the past five years of less than 20 percent; or (b) an enrollment of less than 300 students (TEA, 2015b).”

# Appendix D: Interview Protocol

**\*\*Priority questions are in bold.**

1. Personal background:
   1a. let me confirm your current position….
   1b. Can you tell me how long you've worked here in your position?
   1c. What is your role in evaluating the "community and student engagement performance?

2. Questions about the district:
   2a. Can you tell me about your district demographics, students you serve, and how they've changed in the past five to ten years and/or in recent years?

   2b. Can you tell me about your district general attitude toward House Bill 5 of 2013?
   2c. What aspects of HB 5 does your district prioritize? What is the reason?

3. Perceptions toward the "community and student engagement" (CASE) evaluation:
   3a. When the CASE policy was passed in 2013, how did your district perceive the mandate? (e.g., no guidance from the state)

   3b. What is your district's attitude after two years' implementation (2013-14 & 2014-15)? (e.g. the ratings not comparable across the state; impact)

   3c. In general, does your district see CASE a positive addition or another routine mandate you have to respond?

4. **How did you district determine the rating criteria in the 2013-14 academic year?**
   **4a. Can you tell me about the local district committee in charge of the "community and student engagement" evaluation?**

   **The organization of the committee:** an existing team? a new team?
                              who decided who got onto the committee?
   **Make-up of the committee:** who are on the committee?
   Are campus principals, teachers, parents, local business and community members on the committee? If not, why?
   **Any subcommittee?**
   **What role do you serve?**
   **What is the role of the superintendent during this process?**
   **What role /involvement did the school Board have?**
   What did the committee do? (Set up timeline, calendar, meetings, etc.)

**4b. Can you tell me about the process of determining the rating criteria and evaluation indicators/rubrics?**

      **Which unit was in charge? What role do you play?**

      **Were the criteria internally developed by the district or based on an external model?**

      **How did the deliberating process look like (in the committee and subcommittees)?**

            **Hard to reach consensus?**

            **To whom feedback/output were sought? Did they agree with everything? Did disagreement occur? How it was solved?**

            **How about the revision and finalizing process?**

      **What are the roles of the superintendent and school Board in this process?**

**5. How do you perceive the impact of the "community and student engagement" evaluation?**

      **What do you think has been gained/lost, if anything, in the district?**

      **Are there any unexpected or unintended consequences?**

6. Reflections

      a. What are your district's concerns/challenges, if any? (Regarding time input, associated cost, etc.)

      b. If you could talk to legislators, what would you ask them to change about the "community and student engagement" evaluation?

            (esp., to become part of the accountability system in 2017-18 required by HB 2804 of 2015)

      d. Is there anything else you want to share with me?

# Appendix E: Crosstabulation of Perceptions by District Type

*1. CASE is a necessary supplement to standardized testing scores*

|  |  | independent town | major suburban | urban | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 1 | 1 | 1 | 0 | 3 | 9 | **15** |
|  | % | 4.5% | 2.7% | 3.7% | 0.0% | 4.8% | 10.7% | **5.0%** |
| -2.00 disagree | Count | 5 | 3 | 2 | 3 | 5 | 5 | **23** |
|  | % | 22.7% | 8.1% | 7.4% | 4.4% | 7.9% | 6.0% | **7.6%** |
| -1.00 somewhat disagree | Count | 1 | 1 | 1 | 7 | 5 | 4 | **19** |
|  | % | 4.5% | 2.7% | 3.7% | 10.3% | 7.9% | 4.8% | **6.3%** |
| .00 neutral | Count | 3 | 1 | 2 | 10 | 6 | 5 | **27** |
|  | % | 13.6% | 2.7% | 7.4% | 14.7% | 9.5% | 6.0% | **9.0%** |
| 1.00 somewhat agree | Count | 2 | 8 | 7 | 13 | 9 | 26 | **65** |
|  | % within | 9.1% | 21.6% | 25.9% | 19.1% | 14.3% | 31.0% | **21.6%** |
| 2.00 agree | Count | 8 | 10 | 5 | 23 | 22 | 22 | **90** |
|  | % | 36.4% | 27.0% | 18.5% | 33.8% | 34.9% | 26.2% | **29.9%** |
| 3.00 strongly agree | Count | 2 | 13 | 9 | 12 | 13 | 13 | **62** |
|  | % | 9.1% | 35.1% | 33.3% | 17.6% | 20.6% | 15.5% | **20.6%** |

% within district

*2. CASE offers districts an opportunity to start a conversation*

|  |  | independent town | major suburban | urban | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 0 | 0 | 0 | 1 | 1 | 3 | 5 |
|  | % | 0.0% | 0.0% | 0.0% | 1.5% | 1.6% | 3.6% | 1.7% |
| -2.00 disagree | Count | 2 | 1 | 0 | 0 | 1 | 2 | 6 |
|  | % | 9.1% | 2.7% | 0.0% | 0.0% | 1.6% | 2.4% | 2.0% |
| -1.00 somewhat disagree | Count | 0 | 0 | 0 | 2 | 0 | 4 | 6 |
|  | % | 0.0% | 0.0% | 0.0% | 2.9% | 0.0% | 4.8% | 2.0% |
| .00 neutral | Count | 0 | 1 | 2 | 1 | 4 | 5 | 13 |
|  | % | 0.0% | 2.7% | 7.4% | 1.5% | 6.3% | 6.0% | 4.3% |
| 1.00 somewhat agree | Count | 4 | 8 | 2 | 14 | 19 | 13 | 60 |
|  | % | 18.2% | 21.6% | 7.4% | 20.6% | 30.2% | 15.5% | 19.9% |
| 2.00 agree | Count | 10 | 12 | 11 | 21 | 20 | 32 | 106 |
|  | % | 45.5% | 32.4% | 40.7% | 30.9% | 31.7% | 38.1% | 35.2% |
| 3.00 strongly agree | Count | 6 | 15 | 12 | 29 | 18 | 25 | 105 |
|  | % | 27.3% | 40.5% | 44.4% | 42.6% | 28.6% | 29.8% | 34.9% |

% within district

*3. CASE is a positive addition to the test-based accountability system in Texas*

|  |  | independent town | major suburban | urban | non-metro | other central city | rural | 301 |
|---|---|---|---|---|---|---|---|---|

| | | suburban | | | | | | |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 0 | 1 | 1 | 1 | 3 | 4 | 10 |
| | % | 0.0% | 2.7% | 3.7% | 1.5% | 4.8% | 4.8% | 3.3% |
| -2.00 disagree | Count | 3 | 0 | 0 | 3 | 3 | 4 | 13 |
| | % | 13.6% | 0.0% | 0.0% | 4.4% | 4.8% | 4.8% | 4.3% |
| -1.00 somewhat disagree | Count | 1 | 3 | 1 | 5 | 5 | 6 | 21 |
| | % | 4.5% | 8.1% | 3.7% | 7.4% | 7.9% | 7.1% | 7.0% |
| .00 neutral | Count | 3 | 4 | 2 | 5 | 5 | 5 | 24 |
| | % | 13.6% | 10.8% | 7.4% | 7.4% | 7.9% | 6.0% | 8.0% |
| 1.00 somewhat agree | Count | 1 | 7 | 5 | 13 | 14 | 18 | 58 |
| | % | 4.5% | 18.9% | 18.5% | 19.1% | 22.2% | 21.4% | 19.3% |
| 2.00 agree | Count | 10 | 9 | 11 | 24 | 19 | 25 | 98 |
| | % | 45.5% | 24.3% | 40.7% | 35.3% | 30.2% | 29.8% | 32.6% |
| 3.00 strongly agree | Count | 4 | 13 | 7 | 17 | 14 | 22 | 77 |
| | % | 18.2% | 35.1% | 25.9% | 25.0% | 22.2% | 26.2% | 25.6% |

% within district

### 4. It is important for CASE to be factored into the accountability system in 2017-18

| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 0 | 2 | 0 | 3 | 3 | 5 | 13 |
| | % | 0.0% | 5.4% | 0.0% | 4.4% | 4.8% | 6.0% | 4.3% |
| -2.00 disagree | Count | 2 | 1 | 0 | 2 | 5 | 3 | 13 |
| | % | 9.1% | 2.7% | 0.0% | 2.9% | 7.9% | 3.6% | 4.3% |
| -1.00 somewhat disagree | Count | 3 | 4 | 4 | 4 | 5 | 7 | 27 |
| | % | 13.6% | 10.8% | 14.8% | 5.9% | 7.9% | 8.3% | 9.0% |
| .00 neutral | Count | 2 | 5 | 2 | 9 | 4 | 9 | 31 |
| | % | 9.1% | 13.5% | 7.4% | 13.2% | 6.3% | 10.7% | 10.3% |
| 1.00 somewhat agree | Count | 2 | 7 | 5 | 14 | 14 | 15 | 57 |
| | % | 9.1% | 18.9% | 18.5% | 20.6% | 22.2% | 17.9% | 18.9% |
| 2.00 agree | Count | 9 | 6 | 9 | 21 | 16 | 29 | 90 |
| | % | 40.9% | 16.2% | 33.3% | 30.9% | 25.4% | 34.5% | 29.9% |
| 3.00 strongly agree | Count | 4 | 12 | 7 | 15 | 16 | 16 | 70 |
| | % | 18.2% | 32.4% | 25.9% | 22.1% | 25.4% | 19.0% | 23.3% |

% within district

### 5. CASE under HB5 should have provided local districts some guidance

| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | 300 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 1 | 2 | 0 | 0 | 1 | 2 | 6 |
| | % | 4.5% | 5.4% | 0.0% | 0.0% | 1.6% | 2.4% | 2.0% |
| -2.00 disagree | Count | 2 | 4 | 3 | 3 | 4 | 9 | 25 |
| | % | 9.1% | 10.8% | 11.5% | 4.4% | 6.3% | 10.7% | 8.3% |
| -1.00 | Count | 0 | 1 | 0 | 5 | 3 | 9 | 18 |

| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | |
|---|---|---|---|---|---|---|---|---|
| somewhat disagree | % | 0.0% | 2.7% | 0.0% | 7.4% | 4.8% | 10.7% | 6.0% |
| .00 neutral | Count | 2 | 1 | 1 | 8 | 5 | 8 | 25 |
| | % | 9.1% | 2.7% | 3.8% | 11.8% | 7.9% | 9.5% | 8.3% |
| 1.00 somewhat agree | Count | 3 | 12 | 6 | 15 | 8 | 18 | 62 |
| | % | 13.6% | 32.4% | 23.1% | 22.1% | 12.7% | 21.4% | 20.7% |
| 2.00 agree | Count | 10 | 9 | 10 | 15 | 21 | 21 | 86 |
| | % | 45.5% | 24.3% | 38.5% | 22.1% | 33.3% | 25.0% | 28.7% |
| 3.00 strongly agree | Count | 4 | 8 | 6 | 22 | 21 | 17 | 78 |
| | % | 18.2% | 21.6% | 23.1% | 32.4% | 33.3% | 20.2% | 26.0% |

% within district

## 6. The State should have provided resources

| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | 300 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 1 | 0 | 0 | 0 | 1 | 1 | 3 |
| | % | 4.5% | 0.0% | 0.0% | 0.0% | 1.6% | 1.2% | 1.0% |
| -2.00 disagree | Count | 2 | 4 | 1 | 2 | 5 | 8 | 22 |
| | % | 9.1% | 10.8% | 3.8% | 2.9% | 7.9% | 9.5% | 7.3% |
| -1.00 somewhat disagree | Count | 0 | 2 | 0 | 6 | 1 | 8 | 17 |
| | % | 0.0% | 5.4% | 0.0% | 8.8% | 1.6% | 9.5% | 5.7% |
| .00 neutral | Count | 3 | 3 | 2 | 9 | 7 | 8 | 32 |
| | % | 13.6% | 8.1% | 7.7% | 13.2% | 11.1% | 9.5% | 10.7% |
| 1.00 somewhat agree | Count | 3 | 8 | 5 | 13 | 11 | 14 | 54 |
| | % | 13.6% | 21.6% | 19.2% | 19.1% | 17.5% | 16.7% | 18.0% |
| 2.00 agree | Count | 9 | 10 | 11 | 14 | 14 | 17 | 75 |
| | % | 40.9% | 27.0% | 42.3% | 20.6% | 22.2% | 20.2% | 25.0% |
| 3.00 strongly agree | Count | 4 | 10 | 7 | 24 | 24 | 28 | 97 |
| | % | 18.2% | 27.0% | 26.9% | 35.3% | 38.1% | 33.3% | 32.3% |

% within district

## 7. All the eight rating factors are relevant to my district

| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 0 | 0 | 0 | 1 | 0 | 5 | 6 |
| | % | 0.0% | 0.0% | 0.0% | 1.5% | 0.0% | 6.0% | 2.0% |
| -2.00 disagree | Count | 1 | 2 | 0 | 2 | 1 | 6 | 12 |
| | % | 4.5% | 5.4% | 0.0% | 2.9% | 1.6% | 7.1% | 4.0% |
| -1.00 somewhat disagree | Count | 0 | 1 | 2 | 3 | 2 | 5 | 13 |
| | % | 0.0% | 2.7% | 7.4% | 4.4% | 3.2% | 6.0% | 4.3% |
| .00 neutral | Count | 1 | 0 | 1 | 1 | 3 | 3 | 9 |
| | % | 4.5% | 0.0% | 3.7% | 1.5% | 4.8% | 3.6% | 3.0% |
| 1.00 somewhat | Count | 5 | 3 | 4 | 18 | 12 | 16 | 58 |
| | % | 22.7% | 8.1% | 14.8% | 26.5% | 19.0% | 19.0% | 19.3% |

| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| agree | | | | | | | | |
| 2.00 agree | Count | 9 | 14 | 10 | 31 | 25 | 35 | 124 |
| | % | 40.9% | 37.8% | 37.0% | 45.6% | 39.7% | 41.7% | 41.2% |
| 3.00 strongly agree | Count | 6 | 17 | 10 | 12 | 20 | 14 | 79 |
| | % | 27.3% | 45.9% | 37.0% | 17.6% | 31.7% | 16.7% | 26.2% |

% within district

### 8. A weakness might be that locally evaluated ratings of CASE cannot be compared across the state

| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 11 | 12 | 11 | 24 | 21 | 23 | 102 |
| | % | 50.0% | 32.4% | 40.7% | 35.3% | 33.3% | 27.4% | 33.9% |
| -2.00 disagree | Count | 4 | 12 | 10 | 20 | 22 | 21 | 89 |
| | % | 18.2% | 32.4% | 37.0% | 29.4% | 34.9% | 25.0% | 29.6% |
| -1.00 somewhat disagree | Count | 2 | 4 | 3 | 8 | 4 | 18 | 39 |
| | % | 9.1% | 10.8% | 11.1% | 11.8% | 6.3% | 21.4% | 13.0% |
| .00 neutral | Count | 1 | 1 | 2 | 8 | 6 | 6 | 24 |
| | % | 4.5% | 2.7% | 7.4% | 11.8% | 9.5% | 7.1% | 8.0% |
| 1.00 somewhat agree | Count | 1 | 2 | 0 | 4 | 3 | 3 | 13 |
| | % | 4.5% | 5.4% | 0.0% | 5.9% | 4.8% | 3.6% | 4.3% |
| 2.00 agree | Count | 1 | 3 | 1 | 2 | 5 | 8 | 20 |
| | % | 4.5% | 8.1% | 3.7% | 2.9% | 7.9% | 9.5% | 6.6% |
| 3.00 strongly agree | Count | 2 | 3 | 0 | 2 | 2 | 5 | 14 |
| | % | 9.1% | 8.1% | 0.0% | 2.9% | 3.2% | 6.0% | 4.7% |

% within district

### 9. Since the rating criteria are locally defined, districts should be able to develop their own rating labels
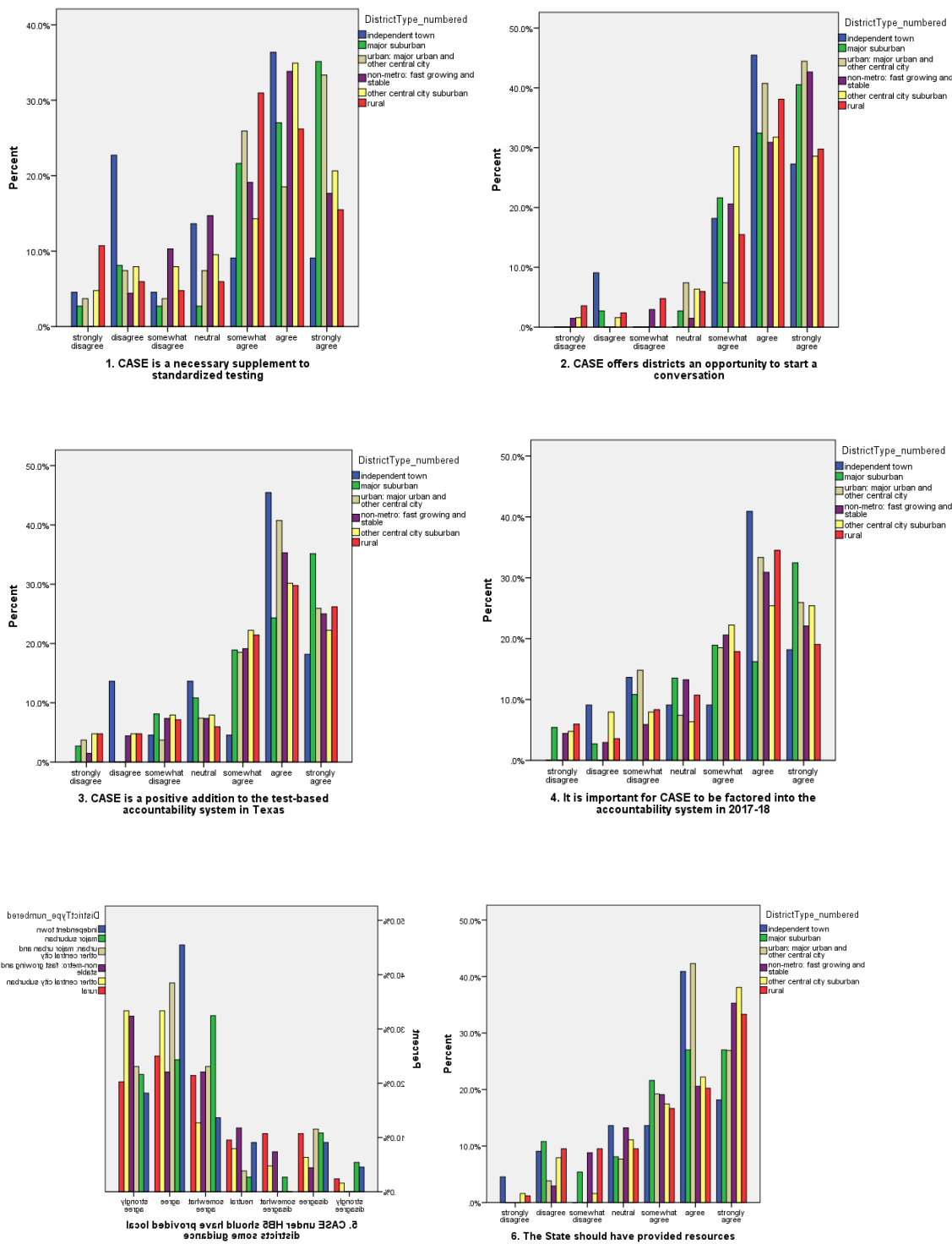
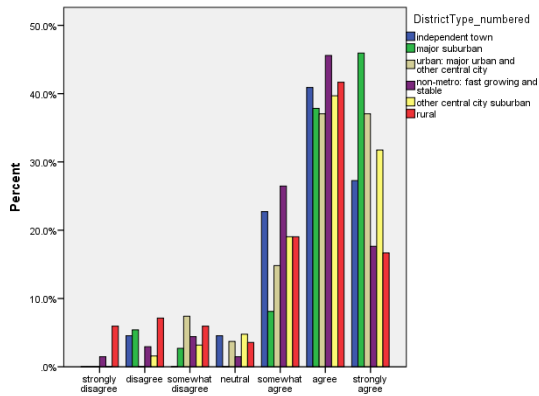| | | independent town | major suburban | urban | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 3 | 4 | 2 | 2 | 2 | 4 | 17 |
| | % | 13.6% | 10.8% | 7.4% | 2.9% | 3.2% | 4.8% | 5.6% |
| -2.00 disagree | Count | 2 | 5 | 7 | 6 | 9 | 8 | 37 |
| | % | 9.1% | 13.5% | 25.9% | 8.8% | 14.3% | 9.5% | 12.3% |
| -1.00 somewhat disagree | Count | 2 | 6 | 4 | 6 | 9 | 4 | 31 |
| | % | 9.1% | 16.2% | 14.8% | 8.8% | 14.3% | 4.8% | 10.3% |
| .00 neutral | Count | 3 | 11 | 2 | 17 | 10 | 18 | 61 |
| | % | 13.6% | 29.7% | 7.4% | 25.0% | 15.9% | 21.4% | 20.3% |
| 1.00 somewhat agree | Count | 0 | 4 | 2 | 7 | 8 | 15 | 36 |
| | % | 0.0% | 10.8% | 7.4% | 10.3% | 12.7% | 17.9% | 12.0% |
| 2.00 agree | Count | 5 | 5 | 2 | 13 | 18 | 19 | 62 |
| | % | 22.7% | 13.5% | 7.4% | 19.1% | 28.6% | 22.6% | 20.6% |
| 3.00 strongly agree | Count | 7 | 2 | 8 | 17 | 7 | 16 | 57 |
| | % | 31.8% | 5.4% | 29.6% | 25.0% | 11.1% | 19.0% | 18.9% |

% within district

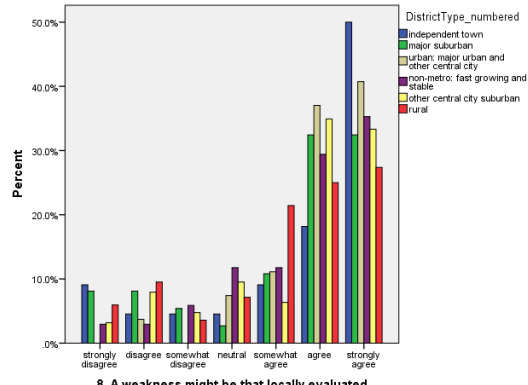## 10. The process of developing the rating criteria of CASE is difficult

| | | independent town | major suburban | urban: | non-metro | other central city suburban | rural | 301 |
|---|---|---|---|---|---|---|---|---|
| -3.00 strongly disagree | Count | 3 | 7 | 4 | 9 | 6 | 12 | 41 |
| | % | 13.6% | 18.9% | 14.8% | 13.2% | 9.5% | 14.3% | 13.6% |
| -2.00 disagree | Count | 4 | 9 | 8 | 15 | 15 | 18 | 69 |
| | % | 18.2% | 24.3% | 29.6% | 22.1% | 23.8% | 21.4% | 22.9% |
| -1.00 somewhat disagree | Count | 7 | 8 | 7 | 17 | 18 | 21 | 78 |
| | % | 31.8% | 21.6% | 25.9% | 25.0% | 28.6% | 25.0% | 25.9% |
| .00 neutral | Count | 1 | 2 | 2 | 7 | 9 | 10 | 31 |
| | % | 4.5% | 5.4% | 7.4% | 10.3% | 14.3% | 11.9% | 10.3% |
| 1.00 somewhat agree | Count | 2 | 3 | 4 | 11 | 7 | 7 | 34 |
| | % | 9.1% | 8.1% | 14.8% | 16.2% | 11.1% | 8.3% | 11.3% |
| 2.00 agree | Count | 2 | 4 | 2 | 6 | 6 | 13 | 33 |
| | % | 9.1% | 10.8% | 7.4% | 8.8% | 9.5% | 15.5% | 11.0% |
| 3.00 strongly agree | Count | 3 | 4 | 0 | 3 | 2 | 3 | 15 |
| | % | 13.6% | 10.8% | 0.0% | 4.4% | 3.2% | 3.6% | 5.0% |

% within district

# Appendix F: More Graphs of Perceptions by District Type



1. CASE is a necessary supplement to standardized testing



2. CASE offers districts an opportunity to start a conversation



3. CASE is a positive addition to the test-based accountability system in Texas



4. It is important for CASE to be factored into the accountability system in 2017-18



5. CASE under HB5 should have provided local districts some guidance



6. The State should have provided resources

7. All the eight rating factors are relevant to my district



8. A weakness might be that locally evaluated ratings of CASE cannot be compared across the state
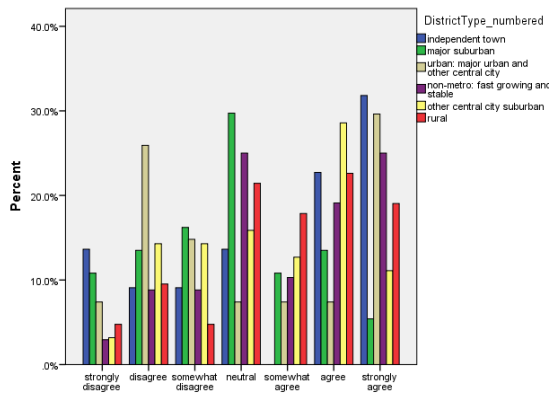


9. Since the rating criteria are locally defined, districts should be able to develop their own rating labels



10. The process of developing the rating criteria of CASE is difficult

174
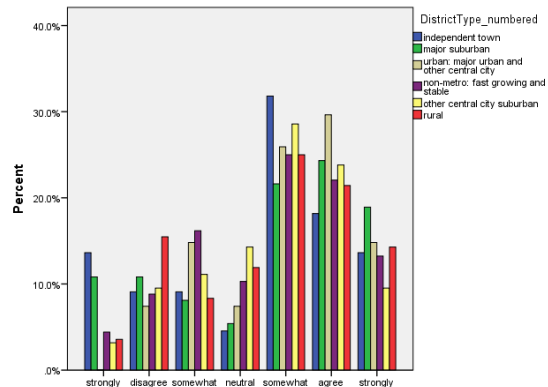
# Appendix G: More on Statistical Results

Tukey Post Doc results: survey question #7, All the eight rating factors are relevant to my district

| (I) District Type | (J) District Type | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| **1.00 independent town** | 2.00 major suburban | -.3084 | .36911 | .961 | -1.3672 | .7505 |
| | 3.00 urban: major urban and other central city | -.1532 | .39377 | .999 | -1.2828 | .9764 |
| | 4.00 non-metro: fast growing and stable | .2139 | .33628 | .988 | -.7507 | 1.1785 |
| | 5.00 other central city suburban | -.1003 | .33952 | 1.000 | -1.0742 | .8737 |
| | 6.00 rural | .6299 | .32836 | .393 | -.3120 | 1.5718 |
| **2.00 major suburban** | 1.00 independent town | .3084 | .36911 | .961 | -.7505 | 1.3672 |
| | 3.00 urban: major urban and other central city | .1552 | .34702 | .998 | -.8403 | 1.1506 |
| | 4.00 non-metro: fast growing and stable | .5223 | .28008 | .426 | -.2812 | 1.3257 |
| | 5.00 other central city suburban | .2081 | .28397 | .978 | -.6065 | 1.0226 |
| | 6.00 rural | .9382* | .27052 | .008 | .1622 | 1.7142 |
| **3.00 urban: major urban and other central city** | 1.00 independent town | .1532 | .39377 | .999 | -.9764 | 1.2828 |
| | 2.00 major suburban | -.1552 | .34702 | .998 | -1.1506 | .8403 |
| | 4.00 non-metro: fast growing and stable | .3671 | .31187 | .848 | -.5275 | 1.2617 |
| | 5.00 other central city suburban | .0529 | .31536 | 1.000 | -.8517 | .9575 |
| | 6.00 rural | .7831 | .30331 | .105 | -.0870 | 1.6531 |
| **4.00 non-metro: fast growing and stable** | 1.00 independent town | -.2139 | .33628 | .988 | -1.1785 | .7507 |
| | 2.00 major suburban | -.5223 | .28008 | .426 | -1.3257 | .2812 |
| | 3.00 urban: major urban and other central city | -.3671 | .31187 | .848 | -1.2617 | .5275 |
| | 5.00 other central city suburban | -.3142 | .23975 | .779 | -1.0019 | .3735 |
| | 6.00 rural | .4160 | .22365 | .429 | -.2256 | 1.0575 |
| **5.00 other central city suburban** | 1.00 independent town | .1003 | .33952 | 1.000 | -.8737 | 1.0742 |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
|  | 2.00 major suburban | -.2081 | .28397 | .978 | -1.0226 | .6065 |
|  | 3.00 urban: major urban and other central city | -.0529 | .31536 | 1.000 | -.9575 | .8517 |
|  | 4.00 non-metro: fast growing and stable | .3142 | .23975 | .779 | -.3735 | 1.0019 |
|  | 6.00 rural | .7302* | .22850 | .019 | .0747 | 1.3856 |
| **6.00 rural** | **1.00 independent town** | **-.6299** | **.32836** | **.393** | **-1.5718** | **.3120** |
|  | **2.00 major suburban** | **-.9382*** | **.27052** | **.008** | **-1.7142** | **-.1622** |
|  | **3.00 urban: major urban and other central city** | **-.7831** | **.30331** | **.105** | **-1.6531** | **.0870** |
|  | **4.00 non-metro: fast growing and stable** | **-.4160** | **.22365** | **.429** | **-1.0575** | **.2256** |
|  | **5.00 other central city suburban** | **-.7302*** | **.22850** | **.019** | **-1.3856** | **-.0747** |

**Based on observed means.**
 **The error term is Mean Square(Error) = 1.880.**
 **\*. The mean difference is significant at the .05 level.**

Tukey Post Doc results: survey question #7, All the eight rating factors are relevant to my district

| (I) District Type | (J) District Type | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| **1.00 independent town** | 2.00 major suburban | .9435 | .49280 | .395 | -.4701 | 2.3571 |
| | 3.00 urban: major urban and other central city | .5051 | .52573 | .930 | -1.0030 | 2.0131 |
| | 4.00 non-metro: fast growing and stable | -.1551 | .44897 | .999 | -1.4430 | 1.1328 |
| | 5.00 other central city suburban | .2193 | .45330 | .997 | -1.0810 | 1.5197 |
| | 6.00 rural | -.0942 | .43839 | 1.000 | -1.3517 | 1.1634 |
| **2.00 major suburban** | **1.00 independent town** | **-.9435** | **.49280** | **.395** | **-2.3571** | **.4701** |
| | **3.00 urban: major urban and other central city** | **-.4384** | **.46331** | **.934** | **-1.7675** | **.8906** |
| | **4.00 non-metro: fast growing and stable** | **-1.0986*** | **.37394** | **.041** | **-2.1712** | **-.0259** |
| | **5.00 other central city suburban** | **-.7242** | **.37913** | **.398** | **-1.8117** | **.3634** |
| | 6.00 rural | -1.0376* | .36117 | .049 | -2.0737 | -.0016 |
| **3.00 urban: major urban and other central city** | 1.00 independent town | -.5051 | .52573 | .930 | -2.0131 | 1.0030 |
| | 2.00 major suburban | .4384 | .46331 | .934 | -.8906 | 1.7675 |
| | 4.00 non-metro: fast growing and stable | -.6601 | .41638 | .609 | -1.8545 | .5343 |
| | 5.00 other central city suburban | -.2857 | .42105 | .984 | -1.4935 | .9221 |
| | 6.00 rural | -.5992 | .40495 | .678 | -1.7608 | .5624 |
| **4.00 non-metro: fast growing and stable** | 1.00 independent town | .1551 | .44897 | .999 | -1.1328 | 1.4430 |
| | 2.00 major suburban | 1.0986* | .37394 | .041 | .0259 | 2.1712 |
| | 3.00 urban: major urban and other central city | .6601 | .41638 | .609 | -.5343 | 1.8545 |
| | 5.00 other central city suburban | .3744 | .32009 | .851 | -.5438 | 1.2926 |
| | 6.00 rural | .0609 | .29860 | 1.000 | -.7956 | .9175 |
| **5.00 other central city suburban** | 1.00 independent town | -.2193 | .45330 | .997 | -1.5197 | 1.0810 |
| | 2.00 major suburban | .7242 | .37913 | .398 | -.3634 | 1.8117 |
| | 3.00 urban: major | .2857 | .42105 | .984 | -.9221 | 1.4935 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | urban and other central city | | | | | |
| | 4.00 non-metro: fast growing and stable | -.3744 | .32009 | .851 | -1.2926 | .5438 |
| | 6.00 rural | -.3135 | .30508 | .908 | -1.1886 | .5616 |
| **6.00 rural** | 1.00 independent town | .0942 | .43839 | 1.000 | -1.1634 | 1.3517 |
| | 2.00 major suburban | 1.0376* | .36117 | .049 | .0016 | 2.0737 |
| | 3.00 urban: major urban and other central city | .5992 | .40495 | .678 | -.5624 | 1.7608 |
| | 4.00 non-metro: fast growing and stable | -.0609 | .29860 | 1.000 | -.9175 | .7956 |
| | 5.00 other central city suburban | .3135 | .30508 | .908 | -.5616 | 1.1886 |

**Based on observed means.**
 **The error term is Mean Square(Error) = 3.351.**
**\*. The mean difference is significant at the .05 level.**

# References

Abedi, J. (2002). Standardized achievement tests and English language learners:

    Psychometrics issues. *Educational Assessment, 8*, 231–257.

Albirini, A. (2006). Teachers' attitudes toward information and communication

    technologies: The case of Syrian EFL teachers. *Computers & Education*, *47*(4),

    373-398.

Alexander, R. (2011). Evidence, rhetoric and collateral damage: The problematic pursuit

    of 'world class' standards. *Cambridge Journal of Education, 41*(3), 265-286.

Allington, R. L., & McGill-Franzen, A. (1992). Unintended effects of educational reform

    in New York. *Educational Policy, 6*(4), 397–414.

Alvarez, L. (2014a, November 9). States listen as parents give rampant testing an F. *The*

    *New York Times*. Retrieved from http://www.nytimes.com/2014/11/10/us/states-

    listen-as-parents-give-rampant-testing-an-f.html?_r=0

Amrein, A. L., & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student

    learning. *Education Policy Analysis Archives, 10*(18). Retrieved from

    http: //epaa.asu.edukpaa/v10n18

Amrein-Beardsley, A. A. & Berliner, D. C. (2003). Re-analysis of NAEP math and

    reading scores in states with and without high-stakes tests: Response to

    Rosenshine. *Education Policy Analysis Archives, 11*(25). Retrieved from

    http://epaa.asu.edu/epaa/v11n25/

Amy, J. (2015, February 12). House votes to eliminate graduation testing requirement.

*Hattiesburg American*. Retrieved from

http://www.hattiesburgamerican.com/story/news/education/2015/02/12/graduatio
n-tests/23313895/

Avramidis, E., Bayliss, P., & Burden, R. (2000). A survey into mainstream teachers'
attitudes towards the inclusion of children with special educational needs in the
ordinary school in one local education authority. *Educational psychology*, *20*(2),
191-211.

Bidwell, A. (2014b, September 2). Florida school district opts out of opting out.
*US News*. Retrieved from

http://www.usnews.com/news/articles/2014/09/02/florida-school-district-retracts-
historic-testing-opt-out-decision

Bidwell, A. (2014a, October 28). Coalition wants new school accountability: Education
accountability should focus more on helping students, teachers and schools
improve. *US News*. Retrieved from

http://www.usnews.com/news/articles/2014/10/28/teachers-unions-education-
advocacy-groups-call-for-new-accountability-system

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas
accountability system. *American Educational Research Journal, 42*(2), 231-268.

Branton, R. (2007). Latino attitudes toward various areas of public policy the importance
of acculturation. *Political Research Quarterly*, *60*(2), 293-303.

Braun, H., Chapman, L., Vezzu, S. (2010). The black-white achievement gap revisited.
*Education Policy Analysis Archives, 18*(21). Retrieved from

http://epaa.asu.edu/ojs/article/view/772

Brennan, R. T., Kim, J., Wenz-Gross, M., & Siperstein, G. N. (2001). The relative

equitability of high-stakes testing versus teacher-assigned grades: An analysis of

the Massachusetts Comprehensive Assessments System (MCAS). *Harvard*

*Educational Review, 71* (2), 173-216.

Brundinnov, J. (2014, November 13). Thousands of students protest Colorado

standardized tests. *Colorado Public Radio*. Retrieved from

http://www.cpr.org/news/story/thousands-students-protest-colorado-standardized-

tests

Burgess, S., Propper, C., Slater, H., Wilson, D. (2005). Who wins and who loses from

school accountability? The distribution of educational gain in English secondary

schools. University of Bristol, Working paper.

Burnette, D. (2016, September 12). ESSA raises K-12 states in 2016 state-level elections.

*Education Week*. Retrieved from

http://www.edweek.org/ew/articles/2016/09/14/essa-raises-k-12-stakes-in-2016-

state-level.html

Camera, L. (2014, July 31). Five states receive NCLB waiver extensions. *Education*

*Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-

12/2014/07/five_states_receive_nclb_waive.html?qs=Five+States+Receive+NCL

B+Waiver+Extensions

Carnoy, M., Loeb, S., & Smith, T. (2001). Do higher state test scores in Texas make for

better high school outcomes? (CPRE Research Report, No. RR-047). Philadelphia,

PA: Consortium for Policy Research in Education, Philadelphia, PA. (ED478984).

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross state analysis. *Educational Evaluation and Policy Analysis, 24*(4), 305-331.

Cavas, B., Cavas, P., Karaoglan, B., & Kisla, T. (2009). A study on science teachers' attitudes toward information and communications technologies in education. *The Turkish Online Journal of Educational Technology*, *8*(2).  Retrieved from http://www.tojet.net/articles/v8i2/822.pdf

Center on Education Policy. (2010). State test score trends through 2007-08, part 6: Has progress been made in raising achievement for English language learners? Washington, DC.

Champion, B. A. (2007). *The effects of high-stakes testing on central office organizational culture: Changes in one school district* (Order No. 3274765). Available from ProQuest Dissertations & Theses Full Text. (304811648). Retrieved from http://ezproxy.lib.utexas.edu/login?url=http://search.proquest.com/docview/30481 1648?accountid=7118

Chang, J. (2015, February 11). Texas, feds on collision course over No Child Left Behind Act. *Statesman*. Retrieved from http://www.statesman.com/news/news/state-regional/texas-education-official-puts-billions-in-federal-/nj8j9/

Chang, Y. F. (2008). Parents' attitudes toward the English education policy in Taiwan. *Asia Pacific Education Review*, *9*(4), 423-435.

Claytor, S. (2015, February 10). More than 300 Albuquerque students opt out of New

Mexico state test so far. *KOB Eyewitness News*. Retrieved from

http://www.kob.com/article/stories/S3703482.shtml?cat=516#.VNtWFOFLUZw

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate

reading policy in their professional communities. *Educational Evaluation and*

*Policy Analysis, 23*(2), 145-170.

Coburn, C. E., & Stein, M. K. (2006). Communities of practice theory and the role of

teacher professional community in policy implementation. In M. I. Honig (Ed.),

*New directions in education policy implementation: Confronting complexity* (pp.

25-46). Albany, NY: State University of New York Press.

Cohen, D. K., & Barnes, C. (1993). Pedagogy and policy. In D. Cohen, M. W.

McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for*

*policy and practice* (pp. 207-239). San Francisco, CA: Jossey-Bass.

Cook, B. G., Semmel, M. I., & Gerber, M. M. (1999). Attitudes of principals and special

education teachers toward the inclusion of students with mild disabilities critical

differences of opinion. *Remedial and special education*, *20*(4), 199-207.

Creswell, J. W. (2003). *Research design: qualitative, quantitative, and mixed methods*

*approaches* (2nd ed.). Beverley Hills, CA: Sage Publications.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods*

*approaches (2nd ed.)*. Thousand Oaks, CA: Sage Publications.

Cruse, K. L. & Twing, J. S. (2000). The history of statewide achievement testing in

Texas. *Applied Measurement in Education*, *13* (4), 327- 331.

Czaja, R., & Blair, J. (2005). *Designing surveys: A guide to decisions and procedures (2nd ed.).* Thousand Oaks, CA: Sage Publications.

Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *Phi Delta Kappa, 73*(3), 220–225.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record, 106*(6), 1047–1085.

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. *Race Ethnicity and Education, 10*(3), 245-260.

Darling-Hammond, L. (2013). What can TALIS tell us? [PowerPoint slides]. Retrieved from https://edpolicy.stanford.edu/sites/default/files/events/materials/ppt-darling-hammond-.pdf

DeMaio, T. J., Rothgeb, J., & Hess, J. (1998). Improving survey quality through pretesting. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 3, 50-58.

Dennis, R. (2000). The role of the state government in public education in the United States. Retrieved from http://www.departments.bucknell.edu/edu/ed370/staterole.html

Denny, D. (2014, March 26). Thousands of parents "opt out" of STAAR testing for their kids. *Kxan*. Retrieved from http://kxan.com/2014/03/26/thousands-of-parents-opt-out-of-staar-testing-for-their-kids/

Denton, D. (n.d.) The Texas reading initiative: Mobilizing resources for literacy. Retrieved from http://www.sreb.org/page/1414/texas_reading_initiative.html

Ding, G., & Hu, X. H. (2004). Xuesheng guozhong keye fudan de genyuan he duice
[Students' heavy academic burden: Causes and solutions]. *Shanghai Education,
21*, 26-29.

Dufour, R. (2016). ESSA: An opportunity for American education. Solution Tree Press.
Retrieved from https://fs24.formsite.com/edweek/images/WP-SolutionTree--
ESSA_An_Opportunity_for_American_Educators.pdf

Dungca, N. (2013, April 22). Cleveland high school students protest standardized testing.
*The Oregonian*. Retrieved from
http://www.oregonlive.com/portland/index.ssf/2013/04/cleveland_high_school_st
udents.html

Dyer, H. S. (1971). The role of evaluation in accountability. Paper presented at
Conference on Educational Accountability, Educational Testing Service,
Washington, D. C.

Education Encyclopedia. (n.d.). Testing-standardized tests and high-stakes assessment.
Retrieved from http://education.stateuniversity.com/pages/2500/Testing-
STANDARDIZED-TESTS-HIGH-STAKES-ASSESSMENT.html

Elementary and Secondary Education Act of 1965 (ESEA of 1965). As Amended
Through P.L. 114-95, Enacted December 10, 2015. Retrieved from
http://www2.ed.gov/documents/essa-act-of-1965.pdf

Elmore, R., & McLaughlin, M. W. (1982). Strategic choice in federal education policy:
The compliance-assistance tradeoff. In A. Lieberman & M. W. McLaughlin

(Eds.), *Policymaking in education, 81st yearbook of the National Society for the Study of Education* (pp. 159- 194). Chicago: University of Chicago Press.

Elmore, R. F. (2002). Unwarranted intrusion. *Education Next, 2* (1), 30-35.

Engdadl, T. (2014, October 31). School board testing discontent rumbles louder as more districts ask for waivers: Two more boards pass resolutions asking for waivers. *Chalkbeat*. Retrieved from http://co.chalkbeat.org/2014/10/31/school-board-testing-discontent-rumbles-louder-as-more-districts-ask-for-waivers/#.VGrh3vnF-Sp

Farkas, G., Hall, L. S., Finn, C. E., Carnine, D., & Meeder, H. (2000). Can Title I attain its goal? *Brookings Papers on Education Policy, 3*, 59-123.

Feller, B., & Hefling K. (2012, February 9). Official: 10 states get ed waiver. *Associated Press*. Retrieved from http://news.yahoo.com/apnewsbreak-official-10-states-ed-waiver-110202341.html

Figlio, D. N., & Getzer, L. S. (2002). *Accountability, ability, and disability: Gaming the system?* Cambridge, MA: National Bureau of Economic Research.

Fink, A. (2009). *How to conduct surveys: A step-by-step guide (4th ed.).* Thousand Oaks, CA: Sage Publications.

Fink, A. (2013). *How to conduct surveys: A step-by-step guide (5th ed.).* Thousand Oaks, CA: Sage Publications.

Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education*. (7th ed.). New York: McGraw-Hill.

Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No

Child Left Behind? *Educational Researcher, 36* (26). doi:

10.3102/0013189X07306556

Gándara, P., & Baca, G. (2008). NCLB and California's English language learners: The

perfect storm. *Lang Policy, 7,* 201-216. doi:10.1007/s10993-008-9097-4

Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in

California schools: Unequal resources, unequal outcomes. *Education Policy*

*Analysis Archives*, *11*(36). Retrieved from http://epaa.asu.edu/epaa/v11n36

Garcia. D.R. (2008). Mixed messages: American Indian achievement before and since the

implementation of No Child Left Behind. *Journal of American Indian Education,*

*47*(1), 136-154.

Garcia, N. (2014, November 10). Two Colorado high schools expect mass opt-outs from

tests this week. *Chalkbeat*. Retrieved from

http://co.chalkbeat.org/2014/11/10/two-colorado-high-schools-expect-mass-opt-

out-of-tests-this-week/#.VGmu9_nF-Sq

Garland, S. (2014, August, 17). Why is a Reagan-era report driving today's education

reform? *The Hechinger Report*. Retrieved from

http://hechingerreport.org/content/report-1980s-driving-todays-education-

reform_17049/

Gravetter, F. J., & Wallnau, L. B. (2008). *Statistics for the behavioral sciences* (8th ed.).

Cengage Learning.

Guisbond, L. (2014). Testing reform victories: The first wave. Retrieved from

http://www.fairtest.org/sites/default/files/TestingReformVictoriesReport.pdf

187

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the

    implementation of instructional innovation. *Teaching and teacher education*, *4*(1),

    63-69.

Guthrie, J. (2011). The update on the STAAR program. Retrieved from

    file:///C:/Users/coura_000/Downloads/STAAR-2011-NovUpdate.pdf

Hagopian, J. (2014, December 31). 2014: The greatest year of revolt against high-stakes

    testing in US history. *Common Dreams*. Retrieved from

    http://www.commondreams.org/views/2014/12/31/2014-greatest-year-revolt-

    against-high-stakes-testing-us-history

Haney, W. (2000).The myth of the Texas miracle in education. *Education Policy Analysis*

    *Archives, 8*(41). Retrieved from http://epaa.asu.edu/epaa/v8n41

Hanushek, E. A. and Raymond, M. E. (2003). Improving educational quality: How best

    to evaluate our schools? In Kodrzycki, Y. (Ed.), *Education in the 21st Century:*

    *Meeting the Challenges of a Changing World* (pp.193-224). Boston, MA: Federal

    Reserve Bank of Boston.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to

    improved student performance?. *Journal of policy analysis and management*,

    *24*(2), 297-327.

Hays, D. G., & Singh, A. A. (2011 or 2012). *Qualitative Inquiry in Clinical and*

    *Educational Settings*. New York, NY: Guilford Press.

Holme, J. J., Carkum, R., & Rangel, V. S. (2013). High pressure reform: Examining

    urban schools' response to multiple school choice policies. *Urban Review, 45*(2),
188

167-196. doi: 10.1007/s11256-012-0216-3

Holme, J. J., Richards, M. P., Jimerson J. B., & Cohen, R.W. (2010). Assessing the effects of high school exit examinations. *Review of Educational Research, 80*(4), 476–526. doi: 10.3102/0034654310383147

Hong, W. P., & Youngs, P. (2008). Does high-stakes testing increase cultural capital among low-income and racial minority students? *Education Policy Analysis Archives, 16*(6). Retrieved from http://epaa.asu.edu/epaa/v16n6/

Honig, M. I. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 1-24). Albany, NY: State University of New York Press.

Hope, M. (2015, January 29). No Child Left Behind waiver at risk in Texas. *Breitbart*. Retrieved from http://www.breitbart.com/texas/2015/01/29/no-child-left-behind-waiver-at-risk-in-texas/

House Research Organization. (2013, December 11). Focus report: Major issues of the 83rd legislature, regular session and first, second, and third called sessions. Retrieved from http://www.hro.house.state.tx.us/pdf/focus/major83.pdf

Howe, K. R., & Meens, D. E. (2012). *Democracy left behind: How recent education reforms undermine local school governance and democratic education*. Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/publication/democracy-left-behind

Jabbar, H. (2014). *The Rising Tide: School Choice and Competition in Post-Katrina New*

*Orleans* (Unpublished doctoral dissertation). University of California, Berkeley.

Jacob, B. A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis, 23*(2), 99-121.

Jacob, B. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics, 89*, 761–796.

Jefferson-Jenkins, C., & Hill, M. H. (2011). Role of federal government in public education:  Historical perspectives. Retrieved from http://lwv.org/files/PES_BP_HistoryofFederalRole.pdf

Jennings, J. L., & Beveridge, A. A. (2009). How does test exemption affect schools' and students' academic performance? *Educational Evaluation and Policy Analysis, 31*(2), 153-175. doi:10.3102/0162373708328468

Johnson, H. L. (2004). Consequences of high-stakes testing: Critical perspectives of teachers and students. *Dissertation Abstracts International*, *65*(08), 2850. (UMI No. 790247951).

Kaestle, C. F., & Smith, M. S. (1982). The federal role in elementary and secondary education, 1940-1980. *Harvard Educational Review, 52*(4), 384-408.

Kantor, H. (1991). Education, social reform, and the state: ESEA and federal education policy in the 1960s. *American Journal of Education, 100*(1), 47-83.

Kantor, H. (2001). In retrospect: David Tyack's the one best system. *Reviews in American History, 29* (2), 319-327.

Keating, Xiaofen. "Sampling Methods and Issues." Class Lecture, Survey Research Methods, The University of Texas at Austin, Austin, October 23, 2014a.

Keating, Xiaofen. "Survey Validity and Reliability Tests." Class Lecture, Survey

    Research Methods, The University of Texas at Austin, Austin, November 6,

    2014b.

Kerchner, C. T. (2014, October 22). California moving rapidly toward post-NCLB

    accountability. *Education Week*. Retrieved from

    http://blogs.edweek.org/edweek/on_california/2014/10/california_moving_rapidly

    _toward_post-

    nclb_accountability.html?qs=California+post+NCLB+accountability

Kim, J., & Herman, J. L. (2009). A three-state study of English learner progress

    (CRESST Report 764). Retrieved from National Center for Research on

    Evaluation, Standards, and Student Testing website:

    http://cse.ucla.edu/products/reports/R764.pdf

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000). *What do test*

    *scores in Texas tell us? Evaluation and Policy Analysis, 8*(49), 1-22.

Klein, A. (2014a, July 21). Waiverless Washington State's request for new NCLB

    flexibility denied. *Education Week*. Retrieved from

    http://blogs.edweek.org/edweek/campaign-k-

    12/2014/07/waiverless_washington_states_r.html

Klein, A. (2014b, July 28). Some states without NCLB waivers say they dodged a bullet.

    *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-

    12/2014/07/some_states_without_no_child_l.html?qs=Some+states+without+NC

    LB+waivers+say+they+dodged+a+bullet

Klein, A. (2014c, November 13). Ed. Dept. to give some waiver states fast track to longer

    renewal. *Education Week*. Retrieved from

    http://blogs.edweek.org/edweek/campaign-k-

    12/2014/11/education_department_will_give.html?qs=Ed.+Dept.+to+Give+Some

    +Waiver+States+Fast+Track+to+Longer+Renewal

Klein, A. (2016, March 31). The Every Student Succeeds Act: An ESSA overview.

    *Education Week*. Retrieved from https://www.edweek.org/ew/issues/every-

    student-succeeds-act/index.html?override=we

Korte, G. (2016, December 11). The Every Student Succeeds Act vs. No Child Left

    Behind: What's changed? *USA Today*. Retrieved from

    http://www.usatoday.com/story/news/politics/2015/12/10/every-student-succeeds-

    act-vs-no-child-left-behind-whats-changed/77088780/

Labaree, D. F. (2000). Resisting educational standards. *Phi Delta Kappa*, *82*(1), 28-33.

Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The

    effects of NCLB's subgroup-specific accountability on student achievement.

    *Educational Evaluation and Policy Analysis, 34*(2), 185-208. doi:

    10.3102/0162373711429989.

Layton, L. (2014, November 24). Oklahoma wins back its No Child Left Behind waiver.

    *The Washington Post*. Retrieved from

    http://www.washingtonpost.com/blogs/local/wp/2014/11/24/oklahoma-wins-

    back-its-no-child-left-behind-waiver/

Lee, J., & Wong, K. (2004). The impact of accountability on racial and socioeconomic

equity: Considering both school resources and achievement outcomes. *American Educational Research Journal, 41*(4), 797–83.

Le Floch, K. C., Martinez, F., O'Day, J., Stecher, B., Taylor, J., & Cook, A. (2007). State and local implementation of the "No Child Left Behind Act." Volume III--Accountability under "NCLB" Interim Report. *US Department of Education*.

Li, J. L. (2002) Jianqing zhongxiao xuesehng guozhong keye fudan wenti de sikao [Pondering on the heavy academic burden problem of primary and elementary school students]. *Journal of Inner Mongolia College of Education, 15*(6), 25-27.

Li, J., & Xu, J. P. (2011). Jianfu yanjiu zongshu [A comprehensive analysis of reducing academic burden]. *Forum on Contemporary Education, 7*, 34-35.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, *29* (2), 4-16.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind act of 2001. *Educational Researcher*, *31*(6), 3-16.

Linton, T. H., & Kester, D. (2003). Exploring the achievement gap between white and minority students in Texas: A comparison of the 1996 and 2000 NAEP and TAAS eighth grade mathematics test results. *Education Policy Analysis Archives, 11*(10). Retrieved from http://epaa.asu.edu/epaa/v11n10/

Liu, L. L. (2013). *Xiaoxuesheng keye fudan yanjiu: jiyu jiaoyu shehuixue shijiao [The study of elementary school students' academic burden: From a perspective of sociology of education]*. (Master's thesis). Retrieved from www.cnki.com.cn

Loeb, S., & Figlio, D. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 3* (pp.383-423). San Diego, CA: North Holland.

Lopez, P., & Valenzuela, A. (2013). House Bill 5: Relating to public school accountability, including assessment, and curriculum requirements. Austin: Texas Center for Education Policy, University of Texas at Austin.

Losen, D., Orfield, G., & Balfanz, R. (2006). *Confronting the graduation rate crisis in Texas.* Cambridge, MA: Harvard University, Civil Rights Project.

Luo, S. F., & Chen, Y. D. (2013). Woguo zhongxiao xuesheng keye fudan wenti yanjiu jinzhang [The research of Chinese primary and middle school students' academic burden]. *Academy, 1*, 85-86.

Malen, B. (2006). Revising policy implementation as a political phenomenon: The case of reconstitution policies. In M.I Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp.83-104). Albany, NY: State University of New York.

Marchant, G. J. & Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives, 13*(6). Retrieved from http://epaa.asu.edu/epaa/v13n6/

Mathis, W. J., & Trujillo, T. M. (Eds.). (2016). *Learning from the federal market-based reforms: Lessons for the Every Student Succeeds Act*. Charlotte, NC: Information Age Publishing.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (2nd ed.).

Thousand Oaks, CA: Sage Publications.

Mauceri (2014). Mixed strategies for improving data quality: The contribution of qualitative procedures to survey research. *Quality and Quantity, 48*, 2773–2790.

Mathews, K. (2014, August 31). One Florida County votes "no more standardized testing". *The Education Cause*. Retrieved from http://www.care2.com/causes/one-florida-county-votes-no-more-standardized-testing.html

McCarty, T. L. (2009). The impact of high-stakes accountability policies on Native American learners: Evidence from research. *Teaching Education, 20*(1), 7-29.

McCauley, L. (2013, April 24). Hundreds of Chicago students walk out of standardized test: "We are over-tested, under-resources and fed up!" *Common Dreams.* Retrieved from http://www.commondreams.org/news/2013/04/24/hundreds-chicago-students-walk-out-standardized-test

McCombs, J. S., Kirby, S. N., Barney, H., Darilek, H., & Magee, S. (2005). *Achieving state and national literacy goals: A long uphill road.* Santa Monica, CA: RAND.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational evaluation and policy analysis*, *9*(2), 171-178.

McLaughlin, M. W., & Pfeifer, R. S. (1988). *Teacher evaluation: Learning for improvement and accountability*. New York: Teachers College Press.

McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In G. Orfield & M.L. Kornhaber (Eds.), *Raising standards or Raising Barriers? Inequity and High-Stakes Testing in Public Education* (pp. 127-150). New York: Century Foundation Press.

195

McNeil, L. M. (2005). Faking equity: High-stakes testing and the education of Latino youth. In A. Valenzuela (Ed.), *Leaving children behind: How "Texas-style" accountability fails Latino youth* (pp. 57-112). Albany, NY: State University of New York Press.

McNeil, L. M., Coppola, E., Radigan, J., & Vasquez Heilig, J. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives, 16*(3). Retrieved from http://epaa.asu.edu/epaa/v16n3/

Megan, K. (2014, December 7). State working on new rating system for schools that goes beyond test scores. *Hartford Courant*. Retrieved from http://www.courant.com/education/hc-school-accountability-1204-20141207-story.html?dssReturn&z=33957

Mehta, J. N., & Kellert, S. R. (1998). Local attitudes toward community-based conservation policy and programmes in Nepal: A case study in the Makalu-Barun conservation Area. *Environmental Conservation*, *25*(04), 320-333.

Menken, K. (2010). NCLB and English language learners: Challenges and consequences. *Theory into Practice, 49*, 121-128.

Merriam, S. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.

Miao, J. & Haney, W. (2004). High school graduation rates: Alternative methods and implications. *Education Policy Analysis Archives, 12*(55). Retrieved from http://epaa.asu.edu/epaa/v12n55/

Mumby, C. (2010, February 7). No pass, no play. No problem? Do school activity

eligibility requirements help students? *The Dagger*. Retrieved from

http://www.daggerpress.com/2010/02/07/no-pass-no-play-no-problem-do-school-

activity-eligibility-requirements-help-students/

Neill, M. (2001). Do high-stakes graduation tests improve learning outcomes? Using

state-level NAEP data to evaluate the effects of mandatory graduation tests. In G.

Orfield & M.L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequity*

*and high-Stakes testing in public education* (pp.107-125). New York: Century

Foundation Press.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing*

*corrupts America's schools.* Cambridge, MA: Harvard Educational Press.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student

achievement: Does accountability pressure increase student learning? *Education*

*Policy Analysis Archives, 14*(1). Retrieved from http: //epaa.asu.edu/epaa/v14n1

Nott, R. (2014, November 17). More parents choosing to opt out of public testing. *The*

*New Mexican*. Retrieved from

http://www.santafenewmexican.com/news/education/more-parents-choosing-to-

opt-out-of-public-school-testing/article_a86a49ee-2ea2-58a5-8233-

fbbcadca3961.html

Orfield, G., Losen, D., Wald, J., & Swanson, C. (2004). *Losing our future: How minority*

*youth are being left behind by the graduation rate crisis*. Cambridge, MA: The

Civil Rights Project at Harvard University.

Ott, A. (2010). A history of high-stakes testing in Texas. Retrieved from

https://classes.lt.unt.edu/Spring_2010/CECS_5420_020/lao0041/Assign%203/ind
ex.html

Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). The consequences of high school
exit examinations for low-performing urban students: Evidence from
Massachusetts. *Educational Evaluation and Policy Analysis, 32*(1), 5-23. doi:
10.3102/0162373709352530

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer,
E. (2004). Methods for testing and evaluating survey questions. *Public opinion
quarterly*, *68*(1), 109-130.

Ravitch, D. (2010). *The death and life of the great American school system: How testing
and choice are undermining education.* New York, NY: Basic Books.

Ravitch, D. (2013, October 28). New Mexico: Parents organize to boycott state tests.
[Web log post]. Retrieved from http://dianeravitch.net/category/chiefs-for-change/

Ravitch, D. (2014a, April 25). Breaking news: Colorado teachers force union to join fight
against high-stakes testing! [Web log post]. Retrieved from
http://dianeravitch.net/2014/04/25/breaking-news-colorado-teachers-force-union-
to-join-fight-against-high-stakes-testing/

Ravitch, D. (2014b, September 11). Good news! A big win against testing frenzy in
Pittsburgh! [Web log post]. Retrieved from
http://dianeravitch.net/2014/09/11/good-news-a-big-win-against-testing-frenzy-
in-pittsburgh/

Ravitch, D. (2014c, November 14). Hundreds of high schoolers boycott new state

standardized test. [Web log post]. Retrieved from

http://dianeravitch.net/2014/11/14/mass-student-opt-out-at-boulders-fairview-

high-school/

Ravitch, D. (2014d, November 14). Breaking news: 97% of seniors at Colorado high

school boycott test. [Web log post]. Retrieved from

http://dianeravitch.net/2014/11/14/breaking-news-97-of-seniors-at-colorado-high-

school-boycott-test/

Ravitch, D. (2014e, November 16). Help Colorado resistance raise $300 for opt out

billboards. [Web log post]. Retrieved from

http://dianeravitch.net/2014/11/16/help-colorado-resistance-raise-700-for-opt-out-

billboards/

Ravitch, D. (2015, January 17). Pennsylvania: Stop using standardized test as graduation

requirement. [Web log post]. Retrieved from

http://dianeravitch.net/2015/01/17/pennsylvania-stop-using-standardized-test-as-

graduation-requirement/

Reardon, S. F., Greenberg, E. H., Kalogrides, D., Shores, K. A., & Valentino,

R. A. (2013). Left behind? The Effect of No Child Left Behind on academic

achievement gaps.

Redmon, J. (2015, February 12). House approves bill to retroactively cancel high school

grad tests. *The Atlanta Journal Constitution*. Retrieved from

http://www.ajc.com/news/news/state-regional-govt-politics/now-in-congress-

georgians-get-a-closer-look-at-wha/nkTtN/

Reid, K. S. (2014, September 26). New York parent group aims to boost number of

students boycotting state tests. Education Week. Retrieved from

http://blogs.edweek.org/edweek/parentsandthepublic/2014/09/new_york_parent_g

roup_aims_to_boost_number_of_students_boycotting_state_tests.html?qs=new+

York+250000+parents+opting+out+of+high+stakes+tests

Reuter, J. (2014, November 10). State tests meet student resistance: Seniors statewide

balk at taking lengthy new assessment. *Lone Tree Voice*. Retrieved form

http://lonetreevoice.net/stories/State-tests-meet-student-resistance,172932

Richburg, J. R. (1971). The movement for accountability in education. Retrieved from

http://files.eric.ed.gov/fulltext/ED058145.pdf

Riderick, M., Bryk, A., Jacob, B., Easton, J., & Allensworth, E. (1999). Ending social

promotion: Results from the first two years. Chicago: Consortium on Chicago

School Research. Retrieved from

https://ccsr.uchicago.edu/sites/default/files/publications/p0g04.pdf

Roberts, R. J. (1995). Can self-reported data accurately describe the prevalence of

overweight? *Public Health*, *109*(4), 275-284.

Rose, L. C., & Gallup, A. M. (2007). The 39th Annual Phi Delta Kappa/Gallup poll of

the public's attitudes toward the public schools. *Phi Delta Kappan*, *89*(1), 33-48.

Rubio, D. M, Berg-Weger, M. Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying

content validity: Conducting a content validity study in social work research.

*Social Work Research, 27*(2), 94-104.

Rumberger, R. W., & Gándara, P. (2004). Seeking equity in the education of

California's English learners. *Teachers College Record, 106*(10), 2032-2056.

Sacramento County Office of Education. (n.d.). Local Control and Accountability Plan (LCAP). Retrieved from http://www.scoe.net/lcap/

Schaeffer, B. (2002, September). Resistance to high stakes testing spreads. *District Administration*. Retrieved from http://www.districtadministration.com/article/resistance-high-stakes-testing-spreads

Serrano, J. (2013, July 19). Leave HB5 to school districts, Aycock tells SBOE. *The Texas Tribune*. Retrieved from http://www.texastribune.org/2013/07/19/aycock-let-local-districts-implement-hb-5/

Shade, R. A., & Stewart, R. (2001). General education and special education preservice teachers' attitudes toward inclusion. *Preventing School Failure: Alternative Education for Children and Youth*, *46*(1), 37-41.

Shaw, L. (2014, July 21). No relief for Washington schools under No Child Left Behind law. [Web log post]. Retrieved from http://blogs.seattletimes.com/today/2014/07/no-relief-for-washington-schools-under-no-child-left-behind-law/

Shuster, K. (2012.) Re-Examining exit exams: New findings from the education longitudinal study of 2002. *Education Policy Analysis Archives, 20*(3). Retrieved from http://epaa.asu.edu/ojs/article/view/797

Sloan, K. (2007). High-stakes accountability, minority youth, and ethnography:

Assessing the multiple effects. *Anthropology and Education Quarterly, 38*(1), 24-41.

Solochek, J. S. (2014, December 2). Florida school boards take a stand against overtesting. *Tampa Bay Times*. Retrieved from http://www.tampabay.com/news/education/k12/florida-school-boards-take-a-stand-against-over-testing/2208695

Solochek, J.S. (2015, February 12). Growing number of parents are seeking testing opt-outs for their children. *Tampa Bay Times*. Retrieved from http://www.tampabay.com/news/education/k12/growing-number-of-parents-are-seeking-testing-opt-outs-for-their-children/2217403

Spillane, J. P. (1996). School districts matter: Local educational authorities and state instructional policy. *Educational Policy, 10(1), 63-87.*

Spillane, J. P. (1998). A cognitive perspective on the role of the local educational agency in implementing instructional policy: Accounting for local variability. *Equational Administration Quarterly, 34*(1), 31-57.

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies, 31*(2), 143-175.

Steele, C. T. (2013, March 16). Students walk out on standardized testing. *Examiner*. Retrieved from http://www.examiner.com/article/students-walk-out-on-standarized-testing

Strauss, V. (2012a, March 15). In Texas, a revolt brews against standardized testing. *The*

*Answer Sheet*. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/post/in-texas-a-revolt-brews-against-standardized-testing/2012/03/15/gIQAI5N0VS_blog.html

Strauss, V. (2012b, November 12). Revolt against high-stakes standardized testing spreads. *The Washington Post*. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2012/11/12/revolt-against-high-stakes-standardized-testing-spreads/

Strauss, V. (2014a, August 28). Florida County opts out of all state-mandated testing in 'act of civil disobedience'. *The Washington Post*. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/08/28/florida-county-opts-out-of-all-state-mandated-testing-in-act-of-civil-disobedience/

Strauss, V. (2014b, October 30). The rise of the anti-standardized testing movement. *The Washington Post*. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/10/30/the-rise-of-the-anti-standardized-testing-movement/

Strauss, V. (2016, January 28). U.S Education Department threatens to sanction states over test opt-outs. *The Washington Post*. https://www.washingtonpost.com/news/answer-sheet/wp/2016/01/28/u-s-education-department-threatens-to-sanction-states-over-test-opt-outs/?utm_term=.0958937c6397

Stutz, T. (2014, July 6). Texas' school reform law of 1984 still touches millions of students. *Dallas News*. Retrieved from http://www.dallasnews.com/news/local-

news/20140706-school-reform-law-of-1984-still-touches-millions-of-texas-students.ece

Tait, K., & Purdie, N. (2000). Attitudes toward disability: Teacher education for inclusive environments in an Australian university. *International Journal of Disability, Development and Education*, *47*(1), 25-38.

Teddie, C., Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications, Inc.

Texas Education Agency. (1994). Accountability manual: The 1994-95 accountability rating system for Texas public schools and school districts. Retrieved from

http://ritter.tea.state.tx.us/perfreport/account/94/manual.pdf

Texas Education Agency. (1997). Texas student assessment program technical digest for the academic year 1996-1997. Austin, Texas: TEA.

Texas Education Agency. (2014a). Student success initiative. Retrieved from http://tea.texas.gov/student.assessment/ssi/

Texas Education Agency. (2014b). STAAR resources. Retrieved from http://tea.texas.gov/student.assessment/staar/

Texas Education Agency. (2015a). District type data. Retrieved from http://tea.texas.gov/acctres/analyze/1314/level.html

Texas Education Agency. (2015b). District type glossary of terms. Retrieved from http://tea.texas.gov/acctres/analyze/1314/gloss1314.html

The Hechinger Report. (2016, June 10). With the Every Student Succeeds Act, have

    policymakers and educators learned from the past? *U.S. News*. Retrieved from

    http://www.usnews.com/news/articles/2016-06-10/with-the-every-student-

    succeeds-act-have-policymakers-and-educators-learned-from-the-past

The Network for Public Education. (2015). Summary of network for public education's

    comments on ESEA draft bill. Retrieved from

    http://www.networkforpubliceducation.org/wp-content/uploads/2015/02/NPE-

    FULL-Statement-on-Discussion-Draft.pdf

Thurlow, M. (2009, Dec 23). Accountability. *Education.com*. Retrieved Nov 3, 2014

    from http://www.education.com/reference/article/accountability/

Turner, C. (2015, December 10). *President Obama signs education law, leaving 'No*

    *Child' Behind. National Public Radio.* Retrieved from

    http://www.npr.org/sections/thetwo-way/2015/12/10/459219774/president-

    obama-signs-education-law-leaving-no-child-behind

Tyack, D. B. (1974). *The one best system: A history of American urban*

    *education.* Cambridge, Mass.: Harvard University Press, 1974.

U.S. Department of Education. (1991). *America 2000: An education strategy:*

    *Sourcebook.* Washington, D.C.: U.S. Department of Education.

U.S. Department of Education. (2004a). Testing: Frequently asked questions. Retrieved

    from http://www2.ed.gov/nclb/accountability/ayp/testing-faq.html#4

U.S. Department of Education. (2004b). Evaluation of Title I accountability systems and

school improvement efforts (TASSIE): First-year findings. Retrieved from

http://www2.ed.gov/rschstat/eval/disadv/tassie1/tassie-short.pdf

U.S. Department of Education (2010). Guidance on school improvement grants under

section 1003(g) of the Elementary and Secondary Education Act of 1965.

Retrieved from http://www2.ed.gov/programs/sif/sigguidance02232011.pdf

U.S. Department of Education. (2015) Every Student Succeeds Act (ESSA). Retrieved

from http://www.ed.gov/essa

U.S. Department of Education (n.d). Turning around the lowest-performing schools.

Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-

flexibility/resources/turn-around.pdf

Valencia, R. R., & Bernal, E. M. (2000). An overview of conflicting opinions in the

TAAS case. *Hispanic Journal of Behavioral Sciences, 2000* (22), 423-444. doi:

10.1177/0739986300224004.

Valenzuela, A. (2004). *Leaving children behind: How "Texas-style" accountability fails

Latino youth* (Ed.). Albany: State University of New York Press.

Valenzuela, A., Fuller, E., & Vasquez Heilig, J. (2006). The disappearance of high school

English learners from Texas high schools. *Williams Review, 1(1)*, 166-200.

Valenzuela, A., Sun, W. L., Germain, E., & Barnes, M. (2015, February 20). High-stakes

accountability in Texas reconsidered [Memorandum]. Austin, Texas: Texas

Center for Education Policy.

Vasquez Heilig, J. & Darling-Hammond, L. (2008). Accountability Texas-style: The

progress and learning of urban minority students in a high-stakes testing context.

*Educational Evaluation and Policy Analysis, 30*(2), 75-110.

doi:10.3102/0162373708317689

Vasquez Heilig, J. (2011). Understanding the interaction between high-stakes graduation tests and English language learners. *Teachers College Record, 113*(12), 1-25.

Vasquez Heilig, J., Young, M., & Williams, A. (2012). At-risk student averse: Risk management and accountability. *Journal of Educational Administration*, *50*(5), 562-585.

Vasquez Heilig, J., Khalifa, M., & Tillman, L. (2013). Why have NCLB and high-stakes reforms failed? Reframing the discourse with a post-colonial lens. In K. Lomotey and R. Milner (Eds.), *Handbook of Urban Education.* (pp. 523-537) New York: Routledge.

Vasquez Heilig, J., Ward, D.R., Weisman, E. & Cole, H. (2014). Community-based School finance and accountability: A New era for local control in education policy? *Urban* Education, *49*(8), 871-894. doi: 10.1177/0042085914557647

Warren, J. R., & Jenkins, K. N. (2005). High school exit examinations and high school dropout in Texas and Florida, 1971-2000. *Sociology of Education, 78,* 122–143.

Warren, J. R., Jenkins, K. N., & Kulick, R.B. (2006). High school exit examinations and state-level completion and GED rates, 1975 through 2002. *Educational Evaluation and Policy Analysis, 28*(2), 131–152.

Weick, K.E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Welner, K. G., & Mathis, W.J. (2015). Reauthorization of the elementary and secondary education act: Time to move beyond test-focused policies. Boulder,

CO: National Education Policy Center. Retrieved from

http://nepc.colorado.edu/files/nepc-policymemo-esea.pdf

Wen, X. Y. (2007). Jianqing xuesheng fudan, huigui jiaoyu benzhi [Reducing students'

academic burden, returning to the essence of education]. *Scientific Consult, 16*,

10-11.

Williams, J. (2006, July 10). Echo chamber: The National Education Association's

campaign against NCLB. *Education Sector*, Retrieved from

http://www.educationsector.org/publications/echo-chamber-national-education-

associations-campaign-against-nclb

Wright, W., & Choi, D. (2006). The impact of language and high-stakes testing policies

on elementary school English language learners in Arizona. *Education Policy*

*Analysis, 14*, 13. Retrieved from http://www.Epaa. asu.edu/epaa/v14n13/

Wright, W. E., & Pu, C. (2005). Academic achievement of English language learners in

post Proposition 203 Arizona (Research Report EPSL-0509-103-LPRU).

Retrieved from http://eric.ed.gov/PDFS/ED508517.pdf

Yin, R. (1981). Life histories of innovations: How new practices become routinized.

*Public Administration Review*, 21-28.

Young, V. (2012). Overview of state assessment program. [PowerPoint slides]. Retrieved

from

http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=214750730

3&libID=2147507294

Zhao, Y. (December 4, 2010). High test scores, low ability. *The New York Times*.

Retrieved from http://www.nytimes.com/roomfordebate/2010/12/02/what-is-a-college-degree-worth-in-china/high-test-scores-low-ability

Zhou, B. (2001). Zhongxiao xuesheng keye fudan guozhong de chengyin jiqi duice [The reason for and solutions to primary and middle school students' heavy academic burden]. *Jilin Education Science, 5*, 19-21.