Summer 2018

A Georgia's School Districts Grades 3-5 Teachers' and Principals' Perceptions Regarding On-Line Georgia Milestones Testing

Doriastino Cheely Brown

A GEORGIA SCHOOL DISTRICT'S GRADES 3-5 TEACHERS' AND PRINCIPALS' PERCEPTIONS REGARDING ON-LINE GEORGIA MILESTONES TESTING

by

DORIASTINO CHEELY BROWN

(Under the Direction of Gregory Chamblee)

ABSTRACT

The purpose of this quantitative research study was to determine a Georgia school district's grades 3 – 5 teachers' and administrators' perceptions of online Georgia Milestones testing on (a) curriculum, (b) teaching, (c) accountability, and (d) stress. Data were collected via survey. Sixty teachers and eight administrators completed the survey. Surveys were analyzed by school, administrator, and domain. Open-ended questions at the end of the survey were analyzed individually by question for teacher and administrator themes. Survey analyses found that teachers and administrators, in the curriculum domain, believed that on-line high-stakes testing led teachers to reassess beliefs about subject matter that is important to teach. The survey analyses of the stress and accountability domains revealed that improving on-line high-stakes scores and accountability grades aided in the stress that teachers experience, however, the presence of on-line testing has increased teachers' awareness of accountability measures. Open-ended question analyses found both teachers and administrators believe that technology glitches and the student's limited keyboarding skills had an effect on the students' overall test scores.

INDEX WORDS: Georgia Milestones, High stakes testing, On-line testing

A GEORGIA SCHOOL DISTRICT'S GRADES 3-5 TEACHERS' AND PRINCIPALS' PERCEPTIONS REGARDING ON-LINE GEORGIA MILESTONES TESTING

by

DORIASTINO CHEELY BROWN

B.S.Ed., University of Georgia, 1977

M.Ed., University of Georgia, 1983

Ed.S., Augusta State University, 1992

A Dissertation Submitted to the Graduate Faculty of Georgia Southern University

in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF EDUCATION

STATESBORO, GEORGIA

A GEORGIA SCHOOL DISTRICT'S GRADES 3-5 TEACHERS' AND PRINCIPALS' PERCEPTIONS REGARDING ON-LINE GEORGIA MILESTONES TESTING

by

DORIASTINO CHEELY BROWN

Major Professor: Gregory Chamblee
Committee: Paul (Mac) Brinson
Sonya S. Gaither

Electronic Version Approved:
July 2018

DEDICATION

This paper is dedicated to my husband, Dave, daughters, Lenore and Ashley, and my brother, Elliott, for continuously encouraging me to continue on this path. To my mother and father, Jacob and Rubye, I say “thank-you” for teaching me when to be persistent and how to be patient in whatever I endeavor to do.

ACKNOWLEDGMENTS

To Dr. Gregory Chamblee, thank you for stepping in to serve as my chair, your untiring assistance and patience to journey with me throughout the whole experience and for providing exceptional advice and help along the way. Without your help, this task would not have been accomplished.

Dr. Sonya Gaither, you have also been with me throughout the entirety of this undertaking. Thank you for continuing to be willing to serve on my committee even after moving away.

Dr. Anne Atkins, to you, it is also with much gratitude that I express to you for coming aboard and working with me to facilitate this process.

To my friend, Vanessa and acquaintance, Kathy, you will never know what it meant to me to receive your kind words of encouragement, guidance and help when my vision became dim.

Dr. Paul (Mac) Brinson, to you, I extend the utmost gratitude for stepping in at the last minute and providing your service.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The No Child Left Behind (NCLB) Act was enacted in 2002. NCLB represented a significant step forward for our nation's children in many respects. It shined a light where students were making progress and where they needed additional support, regardless of race, income, zip code, disability, home language, or background. The law was scheduled for revision in 2007, and, over time, NCLB's testing requirements became increasingly unworkable for schools and educators. Recognizing this fact, in 2010, President Barack Obama joined educators and families calling for the creation of a better law that focused on fully preparing all students for success in college and careers.

In March 2011, in his State of the Union address, President Barack Obama praised efforts to "raise standards for teaching and learning" as well as the federal Race to the Top program which encouraged states, through incentives, to adopt the Common Core State Standards (CCSS). He then called for Congress to pass a bill to overhaul the No Child Left Behind Act before the beginning of the 2011-2012 school year (Cooper, 2011).

On December 10, 2015, President Barack Obama signed the Every Student Succeeds Act (ESSA) into law. This rewrite of the Elementary and Secondary Education Act signaled the start of a new chapter in our country's mission to ensure a high-quality education for all children (Education Trust, 2015). ESSA dramatically reduces the power of the United States Department of Education and gives states authority to design and implement the most appropriate assessments. A goal of ESSA is to ease the pressures of

testing and bring an end to the test prep mania that for years has consumed school officials who worried their schools would not meet Annual Yearly Progress and face various state sanctions (NEA, 2017). ESSA also gives states the ability to set targets for the amount of time spent on testing and to focus on demonstrating student growth.

The Georgia Department of Education introduced the Georgia Milestones Assessment System (Georgia Milestones) statewide during the 2014-2015 school year. Georgia Milestones is a comprehensive summative assessment program spanning grades three through 12. Milestones testing show whether students have the knowledge they need to succeed once they pass to the next grade level. It also gives teachers and parents an ongoing assessment of how well a student is doing and what areas of weakness need to be addressed (Jones, 2014).

The state of Georgia mandated testing would be moved to an on-line format. As a result, student testing will be transitioned to on-line testing over 5 years. In the first year, it is expected that 30 percent of students will take the test on-line. By year 3, it is expected that 80 percent of students will take the assessment on-line. By year 5, the expectation is that 100% of students will be tested on-line.

Fletcher (2011) notes that few schools transitioning to on-line testing will conduct the tests effectively the first time; however, with the proper inventory of technology and proper preparation and planning, the transition to on-line assessments can be effective. In a perfect world, test results are indifferent and should not change no matter what type of test is given whether paper-and-pencil or on-line (Gewirtz, 2013). However, teachers are concerned about testing students in high-needs, urban elementary schools where there has been little practice with on-line test taking and where gaining computer experience in

on-line test taking proved to be one of the greatest challenges (Ogletree, Ogletree and Allen, 2014). Teachers also are concerned that poor testing results will be reflected in their evaluations, Teacher Keys Effectiveness System (TKES) which include student growth and achievement as a significant indicator of teachers' effectiveness. (Owens, 2015). Principals are concerned about recent trends, funding, and accountability. Even with these concerns, it is predictable that the quest to implement computerized testing is inevitable (Schaffhauser, 2011).

Given the level of concern about implementing an on-line testing program in high needs, urban schools, the main purpose of this study was to determine principals and teachers perception of on-line testing as Georgia transitions to Georgia Milestones on-line testing in a non-urban setting. The purpose of this study was to determine grades 3 – 5 teachers' and principals' perceptions regarding Georgia Milestones on-line testing in a Georgia school district.

Statement of Problem

The concept of perception can be unique and arbitrary. Historically, teachers have been tasked with preparing students for testing. Teachers typically administer paper-and-pencil tests in their classrooms during the academic year. Teachers do not typically administer on-line testing of their students during the academic year. With the transition of Georgia Milestones to an on-line format only, in most cases, teachers and principals are now concerned about its impact on (a) curriculum, (b) teaching, (c) accountability, and (d) stress. Research suggests that how teachers implement a policy change such as on-line testing is influenced by their perceptions of that policy (Darling, Hammond, 1990; Honig, 2006). Therefore, it is important to explore teachers' and principals'

perceptions of the Georgia Milestones on-line testing to determine their perceptions of Georgia Milestones on-line testing. Because this issue is in its implementation stage, there was very little research that investigated this question. This study served to begin to fill this gap in the research.

Research Question

How do grades 3-5 principals and teachers perceive Georgia Milestones on-line testing in a Georgia school district?

Significance of the study

This study was significant for several reasons. First, with the on-line learning environment, understanding the perception of on-line testing from the educator's point of view is necessary. Principals need this information in order to design school professional development, develop better test preparation strategies, and find ways to reduce teacher stress. Teachers need this information to decide how to modify their classroom and express to principals their perceptions of this transition.

Second, few research studies exist identifying principals' and teachers' perceptions of on-line testing, especially in Georgia (ERIC search). This study will add to this research area from a district perspective. This information can be used by researchers across the country as comparison data, for example.

Third, this research benefits Superintendents and GADOE by identifying difficulties that are being experienced by school districts in Georgia.

Research Design

The methodological approach for this study is quantitative. A quantitative framework requires that data be specific, concrete, and precise and can be used to assess

the value of on-line testing. Quantitative research offers a systematic scientific analysis of data and their relationships. The quantitative component consisted of administering a survey to principals and teachers to measure their perception of Georgia Milestones on-line testing.

Data Collection

Georgia Southern University Institutional Review Board permission was gained and permission from the Georgia public school system was obtained. A phone call was made to the principals to inform them about the study, to request to engage participants, and explain the extent of teachers and principals' participation. A verbal request was made to initiate a meeting in order to provide participants with the survey with open-ended questions. Surveys were distributed to teachers and principals at a faculty meeting. Completed surveys were put in a legal envelope and placed on a table in the rear of the cafeteria and collected by a designee of the principal and given to the researcher who waited in the front office to insure anonymity. For teachers or principals who were absent, verified by the sign-in sheet, surveys were placed in their boxes with instructions to return the survey in an envelope to the secretary who then placed them in the researcher's box. Surveys were coded by the researcher by colors for each school as well as for teachers and principals. Open-ended questions were analyzed by finding common themes and patterns in questions. Survey data were analyzed by calculating descriptive statistics (mean, standard deviation) by question, school, domain, and administrator.

Limitations

There are none for this study.

Delimitations

1. The findings of this study may not be applicable to other school districts.

2. Findings are limited to concerns of only persons identified as teachers and principals of one school district in Georgia.

Definitions of terms

Georgia Milestones - *A* comprehensive summative assessment program spanning grades 3 through high school. Georgia Milestones measures how well students have learned the knowledge and skills outlined in the state-adopted content standards in language arts, mathematics, science, and social studies (www.gadoe.org).

High stakes testing - High stake tests are tests whose results are used to make decisions that have serious implications for the individuals tested (Lester, 2007, p. 1099).

Assessment system - The combination of multiple assessments into a reporting format that produces comprehensive, credible, dependable information upon which important decisions can be made about students, schools, districts, or states (www.mondofacto.com/facts/dictionary).

Perceptions – Beliefs or opinions held by a group or people (www.meriam-webster.com).

Chapter Summary

The purpose of this study was to determine grades 3 – 5 teachers' and principals' perceptions regarding Georgia Milestones on-line testing in a Georgia school district. Because of the implications on economy and accountability, high sakes testing increases concern among teachers, principals, and the district. After receiving Institutional Review Board approval, school system personnel were contacted, informed of the study, and

permission was requested to perform the study. A survey was administered to the principals and teachers within the same school system. Surveys were collected by a volunteer who gave them to the researcher. Participants were selected using a convenience sample of teachers and principals who had been involved in administering the Georgia Milestones tests. The overall goal was to add value to the field of research in understanding the perceptions of principals and teachers involved when changing high stakes testing that leads from paper-and-pencil to an on-line testing environment.

CHAPTER II

REVIEW OF LITERATURE

Introduction and Background

For decades, schools have relied on traditional test booklets where students completed tests on answer sheets using No. 2 pencils. To some extent, this practice depicted some level of accountability; however, lacked explicit effectiveness. Because the test booklets and the answer sheets were transported in boxes from the testing companies to the schools, massive resources in terms of manpower, money and time, had to be invested in order to ensure the success of the process (Nelson, 2014).

The existence of such problems bundled with the pressures of maintaining sustainability, have eventually led to the need of re-strategizing, which is what has ultimately brought about high stakes testing and such policies as the *No Child Left Behind Act* (NCLB). Testing approaches now are anchored on the need for students at various grade levels taking various tests before graduating to the next grade level by the way of on-line computer systems. By utilizing this innovation, testing companies and assessments to schools using digital platforms have been proven to be more efficient and secure than the paper and pencil (Schaffhauser, 2011). Today, this testing method is being required for use in schools across various states seeing it as a sustainable solution for mapping progress and identifying the weaknesses of each student (Arroya, I., Beal, C., Murray, T., Walles, R., & Woolf, B. 2004). Although this system has been widely successful with various tests in the past, its challenges have also been noted with the elevated amount of concern with the administration of the Georgia Milestones.

In reviewing the landscape of American education over the last 60 years it

becomes obvious that "with each generation of students comes a change in educational practices" (Burks, B.A., Beziat, T.R., Danley, S.,Davis, K.,Lowery, H., & Lucas, J. 2015 p. 253). In 1954, the landmark decision in *Brown v. Board of Education of Topeka* declared the unconstitutionality of racially segregated public education. Two decades later, in 1975, students with disabilities were assured equal opportunities for a free public education under the Individuals with Disabilities Education Act (IDEA). In 2001, the No Child Left Behind Act was passed with the aim of raising educational standards and academic achievement for all students.

Subsequent changes have followed in more rapid succession. The Common Core State Standards (CCSS) began in 2009 as an effort by the National Governors Association Center for Best Practices and Council of Chief State School Officers to create educational standards designed to prepare students for college and careers in a technology-driven global society (Burks et al., 2015; Huddleston & Rockwell, 2015). In 2012, Kentucky became the first state to use the CCSS for their spring term statewide assessment (McLaughlin & Overturf, 2012). With the CCSS still controversial the states began preparing to make the transition from pencil-and-paper testing to on-line assessments aligned with the CCSS (Gewirtz, 2013; Ogletree, Ogletree, & Allen, 2014; Schaffhauser, 2011, 2013). For most of the states and the District of Columbia the target date was spring 2015.

The year 2015 also marked the enactment of the Every Student Succeeds Act the successor to NCLB. Although ESSA, like NCLB is a reauthorization of the Elementary and Secondary Education Act of 1964 (ESEA), and like NCLB, requires statewide grade-level testing and accountability systems, there are some major distinctions. Foremost,

ESSA endows the states with far greater flexibility to determine specifically what is measured and how the measurements are used for accountability purposes (Martin, 2016; Rothman & Marion, 2016). Second, under ESSA the districts are required to consult with stakeholders, including teachers in translating policy into practice (Fennell, 2016). A third distinction is that ESSA eliminates the Adequate Yearly Progress (AYP) requirement, arguably the least popular component of NCLB (Dee & Jacob, 2010; Fennell, 2016; Martin, 2016; Murnane & Papay, 2010).

Dee and Jacob (2010) referred to NCLB as probably "the most far-reaching education policy initiative in the United States over the last four decades" (p. 149). To many educators and administrators, *over-reaching* might be a more apt description. Fennell (2016) noted that educators had numerous nicknames for NCLB and none of them flattering. Rather, they were "expressions of frustrations with the law's arbitrary, unrealistic expectations and its disregard for the expertise of professional educators" (p. 63). Desimone (2013) observed that before NCLB, many teachers embraced standards-based reforms, which inspired them to adopt new teaching strategies to engage their students and improve the quality of teaching and learning. In fact, the late 1980s and early 1990s saw the emergence of innovative modes of assessment such as portfolios and projects (Rothman & Marion, 2016). With the inception of NCLB, the states abandoned these alternatives and reverted to conventional standardized tests.

Guisbond, Neill, and Schaeffer (2012) refer to the era of NCLB as a "lost decade" for educational progress. According to Guisbond et al., as a consequence of its undue reliance on standardized testing, labeling, and punitive sanctions, NCLB not only undermined the education reforms that preceded it, but it also subverted its own goals of

improving academic achievement and neutralizing achievement gaps. Moreover, the threat of sanctions imposed on schools that failed to meet their AYP targets had the most detrimental impact on the disadvantaged students the law was intended to help. Instead of the enriching learning experience the teachers surveyed by Desimone (2013) sought to provide their students, classroom lessons in low-income schools were often little more than preparation for high-stakes standardized tests (Guisbond et al., 2012; Polikoff, 2014).

In a dramatic departure from the previous reauthorizations of the ESEA, ESSA transfers a substantial amount of authority from the federal government back to the states and local education authorities (LEAs; Sharp, 2016). In contrast to the sharp, blanket criticism of NCLB by Guisbond et al. (2012). Boser and Brown (2016) characterize NCLB as well-intentioned but flawed. From their perspective, it was not only a narrow focus on testing that was the problem, but the fact that the multiple-choice tests used for assessments are inadequate measures of student learning and knowledge. They pointed out that in states that espoused standards-based reforms, low-income students showed decisive gains on the National Assessment of Educational Progress (NAEP). However, Guisbond et al. (2012) counter that the students made faster progress before the inception of NCLB. While concurring with Boser and Brown (2016) that fourth grade students experienced gains in mathematics according to NAEP results, Guisbond et al. (2012) noted that between 1996 and 2003, these scores increased by 11 points. Between 2003 and 2011, the math scores rose again but only by six points. Fourth grade reading scores have been essentially stagnant. Several analyses have found that fourth graders' performance in math has improved since NCLB, especially for low-income students, but

there has been no parallel improvement for reading (Dee & Jacob, 2010).

The state of Georgia, the site of this study, adopted the Georgia Milestones Assessment System, which is aligned with the CCSS. The Georgia Department of Education (GADOE, 2017) acknowledged that under the former testing system, the state had some of the lowest achievement standards in the United States. Under the new comprehensive testing system, all students in grades three through 12 are tested in English/Language Arts (ELA) and Mathematics, and students in grades five through 12 are tested in Science and Social Studies as well. High school students take an end of course assessment for 10 courses designated by the GADOE. For a more accurate representation of students' knowledge and skills, the multiple choice questions are augmented with open-ended items on the ELA and Mathematics tests and the ELA assessment includes a writing component.

The Georgia Milestones replaces three individual tests thereby reducing the number of state assessments that the students are compelled to take (GADOE, 2017). For many stakeholders that should be a welcome change. Although standardized tests have been part of American education for decades, the mandates embedded in NCLB elevated the role of testing to an unprecedented degree (Lazarin, 2014). Furthermore, many school districts, especially in urban areas, added assessments beyond the requisite end of year tests. This escalating focus on tests triggered a backlash, with opponents including the nation's second largest teachers union, the American Federation of teachers (AFT; Schaeffer, 2012). Ironically, much of the impetus driving resistance to high stakes testing came from Texas, which was initially in the vanguard of high-stakes assessments. According to the *2013 PDK/Gallup Poll of the American Public,* only22% of respondents

though increased testing had improved the performance of their local schools and an even high proportion (36%) felt testing had the reverse effect on school performance (Darling-Hammond, 2014). Both parents and educators have expressed numerous concerns over excessive testing (Darling-Hammond, 2014; Rothman & Marion, 2016).

Accountability pressures have been implicated in declining job satisfaction among teachers. According to the 2012 *MetLife Survey of the American Teacher,* teachers' job satisfaction decreased by 23 percentage points from 2008, from 62% to 39%, declaring themselves as very satisfied, plummeting to the lowest level in 25 years (Markow, Macia, & Lee, 2013). Roughly half the teachers (51%) reported feeling highly stressed several days a week. Concurrently, educators, counselors, and school psychologists have observed dramatic increases in test anxiety among students in response to increasing emphasis on high stakes assessments (Colwell, 2013; Duffy, Giordano, Farrell, Paneque, & Crump, 2008; Heiser et al., 2015); Segool, Carlson, Goforth, von der Embse, & Barterian, 2013). Studies confirmed that state-mandated high stakes assessments induced greater anxiety than teachers' classroom tests (Heiser, P., Simidian, G., Albert, D., Garruto, J., Faustino, P. & Caci, K., 2015; Segool et al., 2013). Moreover, elementary school students seem to be the most vulnerable to heightened anxiety and stress.

Educators in Georgia are acutely aware of unintended potential consequences of excessive accountability pressures. In 2011, Atlanta was at the center of a cheating scandal involving 178 teachers and principals from 44 public schools (Guisbond et al., 2012). A report by the Georgia Bureau of Investigation (GBI) described a culture marked by "fear, intimidation and retaliation spread throughout the district" (GBI, as cited in Guisbond et al., 2012, p. 5). Additional evidence of cheating was found in Dougherty

County's 11 public schools where teachers reported being coerced into changing their students' test results. Pressure to meet AYP targets was cited as the main reason for cheating.

Most educators, of course, are not involved in unethical practices to produce results. Rather, teachers' intrinsic motivation is damaged by the imposition of mandates that inhibit their autonomy in the classroom and downgrade their professional expertise (Fennell, 2016). In fall 2015, the Georgia Professional Standards Commission reported that close to half (44% to 47%) of the state's public school teachers leave within their first five years of teaching (Owens, 2015). Even for a profession known for high rates of attrition the figure was alarming. In response, the GADOE surveyed more than 53,000 educators to examine potential reasons for this phenomenon. The number and emphasis on mandated tests emerged as the dominant reason for leaving. Frustration with testing was especially prevalent among elementary school teachers.

The second most commonly cited reason for leaving was the state's Teacher Keys Effectiveness System (TKES), which relies on students' achievement and growth as a significant indicator of teachers' effectiveness (Owens, 2015). Many respondents connected the emphasis on high stakes assessments with the TKES, expressing "a sense of injustice that the mandated testing of minors might determine the assessment of an instructor's performance" (p. 4). According to a 2012 report by Scholastic and the Gates Foundation, only 28% of respondents viewed state-mandated assessments as an important measure of student performance and only 26% considered standardized tests to be an accurate portrayal of what students know (Darling-Hammond, 2014).

In measuring student growth the TKES uses a value-added model designed to

estimate a teacher's effect on student achievement based on prior achievement data and other student attributes (GADOE, 2014). Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012) argue that value-added models fail to capture numerous factors affecting students' achievement extraneous to the classroom practices of the individual teacher. The Georgia survey found that the more teachers ascribed attrition to teacher evaluations the less likely they were to recommend teaching as a profession to others (Owens, 2015).

A third recurrent reason for attrition, described in the GADOE report as written "in tones resembling frustration and hopelessness" was the claim that the state decision makers had never spent time "in the classroom" and never solicited input from teachers about changes that would affect them (Owens, 2015, p. 4). The researchers observed a pervasive feeling of being excluded from important decisions that transcended grade level, district, and years of teaching experience.

Fennell (2016) envisions ESSA as an opportunity for teachers to exercise leadership and for "practitioner voice" to be heard "in the development and implementation of education policy" (p. 65). To Darling-Hammond and Falk (2013), both the CCSS and the assessments aligned with them can provide teachers with opportunities for professional learning. From their perspective, "Teacher involvement in the design, use, and scoring of performance assessments has the potential to powerfully link instruction, assessment, student learning, and teacher professional development" (p. 2). The transition from paper-and-pencil to on-line assessment, which is the focus of this study, has the potential to allow more innovative and authentic assessment of learning and provide a more accurate representation of what students actually know (Darling-

Hammond & Falk, 2013; Delen, 2015; Fink, 2015; Rothman & Marion, 2016; Schauffhauser, 2011). At the same time, teachers require professional development matched to the CCSS and to the specific demands of the technology-based tests. More than half the teachers (60%) who served as administrators in field tests of the assessments developed by the Partnership for Assessment of Readiness for College and Careers (PARCC) felt that their on-line training did not fully prepare them for the basic technology issues they encountered (Rennie Center, 2015). Teachers have expressed similar concerns about their preparation for teaching the CCSS (Burks et al., 2015; Hall, Hutchison, & White, 2015).

Hall et al. (2015) explored teachers' perceptions of the CCSS in Writing from the perspective of teachers' pedagogical and philosophical beliefs. According to the researchers, it is important to examine teachers' perceptions of a new education reform for two reasons. First, if teachers sense a lack of support from their districts and are not satisfied with the professional development they receive in relation to the CCSS it is less probable that they will maintain the high level of expectations embedded in the standards. Second, teachers are unlikely to maintain fidelity to the standards if they perceive them as incompatible with their personal theories and beliefs about teaching and learning, including their beliefs about what constitutes appropriate expectations for students.

Desimone (2013) observed that in the pre-NCLB reform era, teachers framed adherence to standards according to their beliefs about how and what their students should learn. While some teachers made minimal adjustments, such as simply changing the order in which subjects were taught, others saw opportunities to cover new material, place more emphasis on conceptual understanding than on correct answers, and present

lessons in ways that enhanced students' comprehension and retention. Similarly, teachers who already feel they are under duress due to high-stakes assessments may view the transition to online assessment as an added burden, while others may shared the perspective that new modes of assessment can mean new opportunities for teachers to become active in innovation and leadership. Only through research exploration is it possible to understand how teachers really perceive education reforms that affect them. With this knowledge teachers can be supported in accordance with their specific needs and preferences.

On-line high-stakes testing is not actually novel. The state of Virginia has been conducting online assessments since the early 2000s (Schauffhauser, 2011). By the 2005-2006 academic year 21 states and the District of Columbia offered on-line testing (Kim & Huynh, 2007). Historically, the transition from paper-and-pencil to on-line tests can be viewed as the most recent advancement in the long history of standardized testing in the U.S. The following section will discuss the evolution of high stakes assessments in the American public education system.

Federal Mandates

Rothman and Marion (2016) observed that "Assessment has long had a prominent—and controversial—role in American education" (p. 34). Until the mid-19th century the primary mode of assessment for American schoolchildren was oral recitation (Huddleston & Rockwell, 2015). The first major shift occurred in 1845 when Horace Mann argued in favor of having students take standardized written exams. The public education system was expanding rapidly and advocates of written exams believed they had the capacity to evaluate the growing numbers of students more objectively and

efficiently than oral recitations.

The early 20th century saw the development of the standardized IQ and achievement tests that have left an indelible (and to critics, dubious) legacy (Huddleston & Rockwell, 2015). In 1905, psychologists Alfred Binet and Theodore Simon created the first IQ test, designed to identify French students deemed incapable of succeeding in school. Their innovation was followed in 1914 by the development of an American version, the Stanford-Binet test, by psychologists Henry Goddard, Edmund Huey, and Lewis Terman. During the same time, Edward Thorndyke and his students at Columbia University were involved in developing achievement tests in arithmetic, reading, spelling, language, drawing, and handwriting. Although the idea of standardized handwriting tests may seem quaint today, Thorndyke's handwriting test was one of the first norm-referenced assessments and had a powerful impact on the subsequent development of standardized academic achievement tests.

Whereas Horace Mann's compelling interest in education arose from the belief that all children could learn with good teaching, the poor test performance of an expanding and increasingly diverse public school population led educators to feel more aligned with the intentions of Binet and Simon to distinguish between more and less capable students (Clarke, Madaus, Horn, & Ramos, 2000). IQ testing allowed the educators to divert blame for poor academic performance from the schools to the students' inherent abilities. "Scientific" intelligence and achievement tests were used to create the ability tracking system used to place students in the "proper" curriculum track throughout the 20th century (Clarke et al., 2000, p. 162). This was actually a delayed consequence of the early tests, which were too resource-intensive to be used on any large

scale.

During World War I, however, group intelligence tests were developed as a mechanism for classifying recruits (Clarke et al, 2000; Huddleston & Rockwell, 2015). By 1932 (not coincidentally the era of the eugenics movement), three-quarters of large city school systems in the U.S. were using group intelligence tests for tracking purposes and some college employed them to rationalize their admission decisions (Clarke et al.., 2000).

The pivotal factor in the evolution of large-scale testing was the creation of the multiple-choice format by Frederick J. Kelly in 1914 (Clarke et al., 2000). However, for widespread dissemination the tests still had a serious drawback: they had to be graded by hand using pencils, a process not only time-consuming but subject to human error as well (Huddleston & Rockwell, 2015). That situation was changed by a now obsolete gadget called the Markograph, the creation of a high school science teacher. The Markograph was capable of electronically sensing if pencil marks indicated the correct answer. Johnson sold the Markograph to IBM, which produced a number of rapid test-scoring machines throughout the 1930s and 1940s. A decade later, in 1955, the invention of the high-speed scanner simultaneously reduced costs and heightened efficiency, making it possible to administer standardized multiple-choice tests to virtually every student across the U.S. (Clarke et al., 2000).

The invention of a device that enabled large-scale testing occurred at the height of the Cold War. In the aftermath of the Soviet launch of the satellite Sputnik in 1957, Congress passed the National Defense Education Act of 1958, which was followed in 1965 by ESEA (Huddleston & Rockwell, 2015). Both legislative acts mandated

expansion of standardized testing. During the 1970s evidence of declining test scores led to renewed emphasis on the mastery of basic skills and calls for more rigorous school curriculum. However, the driving force in the reform movements that swept through the schools in the 1980s and 1990s was the 1983 publication of *A Nation at Risk,* the scathing critique of the U.S. educational system by the National Commission on Excellence in Education (Boser & Brown, 2016; Huddleston & Rockwell, 2015).

A Nation at Risk called for major changes in five areas: standards, content, time, teachers, and leadership (Boser & Brown, 2016). Specific recommendations included higher standards for colleges and universities, more instructional time devoted to core academic subjects, regular assessment of mastery of content, a professionalized teaching force, and special attention to the needs of socially and economically disadvantaged groups. Momentum for education reform escalated over the next decades, leading to first to the Goals 2000: Educate America Act signed into law by President Bill Clinton and finally to the reauthorization of the ESEA that became NCLB. Even before the enactment of NCLB one of the effects of *A Nation at Risk* was a massive expansion of standardized testing (Huddleston & Rockwell, 2015).

High stakes testing can be broadly defined as any assessments that carry "direct and significant consequences for the person or institution being tested or assessed" (Duffy et al., 2008, p. 53). The IQ and achievement tests of the early and mid-20th century unquestionably classify as high stakes testing, as they decided the future direction of a student's education, or whether it should be continued at all. In the 21st century, however, the term high stakes testing is typically preceded by *mandated,* meaning "the tests and assessments are required by local, state, and/or federal authority" (Duffy et al.,

2008, p. 53). In the minds of educational stakeholders, the term *mandated high stakes testing* is often equated with NCLB.

As previously stated, the component of NCLB that drew the greatest degree of criticism was the AYP whereby schools could be sanctioned for failing to meet yearly progress goals. Critics argued that this burden fell heavily on the schools serving the disadvantaged students the law was intended to help (Guisbond et al., 2012). At the other end of the socio-demographic spectrum, the AYP requirement also hurt teachers, administrators, and students at high-performing schools where a ceiling effect limited the extent that students were capable of improving. With respect to classroom practices, the most common criticism is that teachers were forced to sacrifice enriching activities in favor of a narrow focus on the academic subjects being tested, specifically reading and mathematics. The term "teaching to test" became common usage. Another term that came into being was "bubble kids," denoting students who were close to meeting proficiency standards (Murnane & Papay, 2010). Because AYP was based on the percentage of students from each subgroup who met the proficiency level, teachers were pressured to focus attention on the "bubble kids" at the expense of the high-achievers who would meet the proficiency standards without extra help and the struggling students who would still fall short of proficiency even with individualized attention.

According to Murnane and Papay (2010), probably the most prevalent criticism of the AYP requirements is they did not provide a valid measure of the schools' success in improving the knowledge and skills of students. Teachers offered a number of reasons for this misalignment. First, the rules classify a school as "failed to meet AYP" if it did not meet the standard for any single subgroup. Second, the AYP formula did not reward

significant gains in the achievement of very low-performing students unless they (improbably) managed to meet the proficiency level. Third, teachers complained that the formula did not consider the situation of schools with high concentrations of children who were poor, had special needs, or had limited English proficiency, which made it more difficult to reach AYP targets. Paradoxically a law that was supposed to narrow (or ideally, eradicate) academic achievement gaps worked against the best interests of many students who were already disadvantaged.

As of August 1, 2016, the AYP provision was discarded (Fennell, 2016). Drawing on national data, including NAEP data and RAND surveys, Dee and Jacob (2010) reported some benefits of high stakes testing mandated by NCLB although they noted negative effects as well. A significant finding was an improvement in elementary school mathematics which was most apparent among Hispanic students. However, there was no comparable improvement in reading performance at any grade level or for any subgroup of students. The RAND findings suggested that educators were using formative or diagnostic assessments to gauge students' strengths and weaknesses and focus instruction accordingly. Teachers also reported increases in technical assistance and professional development opportunities. At the same time there was evidence that under pressure to boost achievement, teachers were narrowing the curriculum and teaching to the test. Some evidence suggested an estimate increase in "behavioral engagement," which was twice as pronounced in high-poverty schools. However, Dee and Jacob (2010) acknowledged that their own analysis in another published study disputed the magnitude of that effect.

At the time Dee and Jacob's (2010) report was published, ESEA had not yet been

reauthorized but the Obama administration had released a "blueprint" outlining certain prospective features, notably a continuation of annual school-level summative assessment but allowing the states flexibility in how they calculated school effectiveness. In addition, the proposal called for the use of non-test accountability measures, in particular indicators of college readiness. Another prospective feature was a greater degree of flexibility in how the states decided to intervene in the situation of the most underperforming schools and schools with consistently large achievement gaps. Although Dee and Jacob were somewhat skeptical of "state reforms that decouple performance measures from meaningful consequences" (p. 191), the greater flexibility and elimination of AYP requirements in the reauthorized ESEA were largely welcomed in the education community (Fennell, 2016; Martin, 2016; Rothman & Marion, 2016; Sharp, 2016).

Common Core State Standards

According to Boser and Brown (2016), like its predecessors in standards-based reforms, Common Core is a product of "the shared desire of educators, policymakers, and other stakeholders to raise academic standards in schools" (p. 6). What distinguishes the CCSS is that the standards are being adopted by multiple states, along with high-quality aligned assessments for all students in their K-12 education system. In addition, local school districts are playing a more prominent role in the adoption of high-quality curricula and targeted capacity building. In a parallel fashion, what distinguishes the transition to online testing is not the use of technology-based testing, which some states have been doing for several years but that virtually all school districts are switching to on-line testing within an established time frame (Schauffhauser, 2011, 2013).

Beyond the scope of implementation, the CCSS are also distinguished by the principles underpinning the standards. Specifically, the CCSS: (a) are aligned with college and work expectations; (b) are clear, understandable, and consistent; (c) include rigorous content and application of knowledge via higher order skills; (d) build upon strengths and lessons of current state standards; (e) are informed by other top-performing nations so that all students are prepared to succeed in a global economy and society; and (f) are evidence-based (McLaughlin & Overturf, 2012, p. 153). According to McLaughlin and Overturf, the guiding philosophy of the CCSS is reflected in the College and Career Readiness Standards. Students who can be classified as college and career readiness possess a number of important attributes including independence, strong content knowledge, the ability to respond to varying the varying demands of audience, task, purpose, and discipline, the ability to comprehend was well as critique, appreciation for evidence, competence using technology and digital media strategically and capably, and understanding of other perspectives and cultures.

Huddleston and Rockwell (2015) argue that the emphasis on high stakes testing culminating in NCLB grew out of a manufactured educational crisis. Although the Center for American Progress (CAP) is critical of the excessive emphasis on testing (Lazarin, 2014) as well as reliance on multiple-choice tests that ignore higher order cognitive skills, the CAP has declared that “The plain fact is that the U.S. education crisis is real” (Boser, U., Balfour, P., & Vela, S., 2016, p. 13). CAP has enthusiastically endorsed the CCSS as part of a long-range strategy to “promote equity and excellence among all students” (p. 13).

To CAP the problem is not declining or stagnant test scores but rather the

existence of "a massive divide between what the K-12 system expects of students and what the world of work expects of the nation's students" (Boser et al., 2016, p. 13). They envision the CCSS as a mechanism for bridging this chasm as well as increasing the level of rigor in public schools. Unlike prior attempts to boost academic achievement that resulted in narrowing the curriculum and focusing mainly on mastery of basic skills, the CCSS will help students develop the critical thinking skills that are essential to college and career success. Furthermore, they strongly believe that "when it comes to addressing the nation's education crisis, there is a clear consensus that higher standards can help drive up achievement" (p. 13). In the decade before NCLB that was indeed the perspective of teachers who were inspired by standards-based reforms to seek out creative ways to enhance the quality of teaching and learning in their classrooms (Desimone, 2013).

The states have embraced the standards for two key reasons. First, adopting a "common set of standards" was requisite for states applying for federal Race to the Top funds (McLaughlin & Overturf, 2012). Second, even after adopting the CCSS individual states have the flexibility to modify them by adding up to 15% of new content. Georgia is one of the states involved in the Race to the Top (GADOE, 2014). Georgia's state reform agenda, supported by a Race to the Top grant sets five objectives: (1) Set high standards and rigorous assessments for all students, leading to college and career readiness; (2) Prepare students for college, transition, and success; (3) Provide great teachers and leaders; (4) Provide effective support for all schools, including the lowest-performing schools; and (5) Lead the way in science, technology, engineering, and mathematics (STEM) fields. Notably, one of the challenges cited in the year 3 report is

the implementation of the state's teacher and leader evaluation systems (GADOE, 2014), which has been met with apprehension by Georgia public school teachers (Owens, 2015).

Most educators have positive attitudes toward the CCSS (Markow et al., 2012). However, the standards are designed to be multifaceted, which can make using them a complex endeavor (McLaughlin & Overturf, 2012). In order to use the standards to maintain high-quality teaching and learning, teachers require appropriate professional development and support (Burks et al., 2015; Hall et al., 2015; McLaughlin & Overturf, 2012). Not all teachers appear to be getting the support and assistance they need. Soliciting input from teachers on the CCSS and the companion assessments will help the schools optimize this opportunity to replace previous ineffective standards and related assessments with a system that will fully address the diverse needs of learners in the state's public schools.

McLaughlin and Overturf (2012) observed that elementary school teachers who have adopted the ELA standards "find that their thought processes about curriculum, instruction, and assessment are being continually challenged" (p. 157). In contrast to the narrowing of the curriculum under NCLB, classroom teachers have a substantial amount of flexibility in using the CCSS. This can provide them with opportunities to be creative in providing their students with enriching learning experiences as many teachers did during the standards-based reforms two decades ago (Desimone, 2013). At the same time, the CCSS carries greater measures for accountability. McLaughlin and Overturf (2012) noted that many teachers will discover that they need to alter their classroom practices in order to help their students meet the high expectations embedded within the CCSS.

Burks et al. (2015) explored middle and high school teachers' perceptions of implementing the CCSS with regard to both their feelings about adapting to changes and the question of whether they had received sufficient training and preparation. The participants were 35 teachers from schools in Alabama, Texas, South Carolina, and Maryland. Most (71%) had at least seven years of teaching experience, thus they could directly compare their experience with Common Core with the state's previous standards. More than half the teachers (57%) reported feeling "comfortable" or "extremely comfortable" implementing the CCSS compared to 26% who felt "uncomfortable" or "extremely uncomfortable." The remaining teachers were neutral. Interestingly, it was the more experienced teachers who expressed neutral or negative feelings toward implementing the CCSS while the overwhelming majority (80%) of teachers with less than seven years of experience, including novices, felt at ease with the new standards.

The amount of training they received for the purpose of implementing the CCSS was a concern for most teachers (Burks et al., 2015). More than half the teachers felt their preparation had been inadequate although a substantial proportion (47%) had been involved in three or more related professional development activities. Most of the training occurred at the schools were the teachers taught. Burks et al. noted that this seems to be common practice; most professional development takes place at the school site led by district staff.

Although the sample was small, Burks et al. (2015) noted that their findings were findings from a national survey by the EPE (Editorial Projects in Education) Research Center. The EPE survey also questioned the teachers on whether they felt their students were prepared to master the standards. Only about one-quarter (24%) felt their students

were ready for the new, more challenging standards. Absent from both studies was *why* the teachers felt as they did for reasons other than training. The fact that more experienced teachers were less likely to feel comfortable with the CCSS raises the question of whether their experiences with previous standards might have left them feeling cynical about the new standards or whether they might have been less amenable to changing their teaching their teaching practices in response. The fact that some teachers felt insufficiently prepared despite having been involved in professional development activities also raises questions about the quality of the professional development and training teachers are provided by local schools districts.

Hall et al. (2015) explored teachers' perceptions of the Common Core standards in Writing. Teachers from Georgia were among the 250 K-12 teachers, along with educators from Kentucky, Minnesota, Michigan, North Carolina, Maine, Mississippi, and Delaware. All of these were states that had fully implemented the standards by spring 2013. As in the study of Burks et al. (2015), the teachers were asked about their preparation and professional development related to the CCSS, as well as their perceptions of barriers to implementation and positive and negative effects of the CCSS.

Hall et al. (2015) prefaced their research with quotes from two teachers reflecting very disparate view of the Common Core writing standards. To one teacher the CCSS in Writing were merely "an upper middle class pipe dream about education" that was not relevant to the lives of her students (p. 88). The other teacher lauded the CCSS in Writing for increasing "the value of effective communication for students and teachers," and for their "focus on real world applications" and ability to "reduce reliance on textbooks and irrelevant literature, and better prepare students for the next level" (p. 88).

Hall et al. noted that teachers' educational practices and interactions with students are influenced by their personal philosophies of teaching and learning. They also acknowledged that implementation of the CCSS is affected by external forces such as funding which can be scarce. One of the criticisms of NCLB is that teachers were often not supported in their efforts to improve academic performance (Boser & Brown, 2016). According to Hall et al. (2015), in an era of funding constraints the states need to be creative in finding ways to support teachers in adopting the CCSS.

Hall et al. (2015) drew on teachers' pedagogical and philosophical beliefs as the framework for their on-line survey. Most of the teachers reported feeling "somewhat" familiar with and prepared to teach the CCSS in Writing. Although the study did not explore their reasons, Hall et al. noted that the teachers favored opportunities to "break down" the standards due to their complexity. It was also evident that teachers who received more professional development felt more prepared to adopt the standards. Primary grade teachers (K-3) felt most at ease with and prepared to teach the standards, suggesting that ELA middle and high school warrant more professional development and support. In contrast to Burks et al. (2015), who found less experienced teachers more comfortable with the standards, teachers with less classroom experience felt less prepared to teach the CCSS in writing. Hall et al. (2015) attribute this to the fact that most few teacher education programs require candidates to take methods courses on teaching writing.

Lack of time and resources emerged as the major barriers to implementing the CCSS in Writing. Although lack of time and resources are consistently cited as obstacles to educational innovations, Hall et al. (2015) pointed out that teachers have always had to

scramble to find sufficient time to teach writing. One excellent strategy for giving higher priority to writing would be greater integration of writing into all subject areas at all grade levels. Nearly three-quarters (72%) of the teachers felt that they would be more confident teaching writing if they had more curricular materials aligned with the CCSS. Many teachers thought it would be helpful to watch videos of teachers implementing the standards, along with live demonstrations by experts and colleagues, and/or samples of students' work and assessments aligned with each standard. These techniques would be beneficial for building self-efficacy (Bandura, 1997).

The findings revealed that many teachers had disparate opinions regarding the value and feasibility of the CCSS in Writing (Hall et al., 2015). Nevertheless, the heightened rigor and high expectations were almost universally regarded as positive result. While most teachers welcomed the increased rigor, many also expressed concerns about the developmental appropriateness for individual standards, with concurrent concerns about how to differentiate instruction or adjust the pace of their lessons. A number of teachers felt that their students lacked adequate background knowledge for meeting the more demanding performance standards. Teachers are called on to help students bridge gaps in knowledge and skills resulting from the contrast between the way subjects were previously taught and the standards the students are now expected to meet.

Some teachers expressed the belief that the CCSS in Writing are not appropriate for all students, especially English language learners (ELLs) and students performing below grade level (Hall et al., 2015). The teachers felt least prepared to teach the standards to ELLs and students with intellectual disabilities, which Hall et al. noted is a common concern for teachers, even many veteran teachers. All teachers should be

provided with professional development opportunities that enable them to meet the needs of diverse learners (Darling-Hammond & Falk, 2013). Darling-Hammond has consistently expressed support for communities of practice where teachers share ideas and engage in collaborative learning.

Notably, more than half the teachers (54%) received one day or less of professional development related to the standards (Hall et al., 2015). Teachers surveyed by Burks et al. (2015) received more professional preparation but still felt insufficiently prepared. McLaughlin and Overturf (2012) argue that in order to use the CCSS effectively, teachers need to begin with a Common Core Implementation Plan in addition to having access to numerous professional development opportunities at the state, regional, and district levels. At the district level they assert that in order to lead CCSS initiatives, district staff needs to adopt a proactive stance to supporting teachers and principals, securing material and financial resources, ensuring support for curriculum alignment, and leading professional development activities.

As the first state to adopt the CCSS, Kentucky has an array of strategies and supports to assist teachers, including a summer institute (McLaughlin & Overturf, 2012). One of the problems identified by teachers attending the summer institute was the type of gap in students' conceptual understanding that concerned the teachers about the CCSS in Writing (Hall et al., 2015). To meet the needs of their students, teams of teachers designed formative assessments including questions to informally assess the knowledge and ability of individual students to work with the standards. At the institute, the teachers formed the professional learning communities envisioned by Darling-Hammond and Falk (2013).

Schoenfeld (2015) advocates using formative assessments as a strategy for helping students cultivate the knowledge and skills needed to meet the CCSS in mathematics. Although the National Council of Teachers of Mathematics (NCTM) developed standards that emphasize problem solving, analytical thinking, and real world applications of knowledge more than 25 years ago, the focus of many teachers remained on procedural knowledge. According to Schoenfeld, implementing the CCSS in mathematics can be particularly challenging for teachers whose teacher had a procedural orientation. Formative assessments can be used as a diagnostic tool to discover students' strengths and weaknesses for learning aligned with the standards and can also be used to prepare students for the annual school-based summative assessment.

In light of the transition to on-line testing, on-line formative assessments can be used to familiarize students with technology-based tests (Cassady & Gridley, 2005). A particular advantage of computerized formative assessment is that learners can be provided with immediate feedback. In fact, the provision of feedback has been considered one of the benefits of computer-assisted learning since its inception in the 1970s. A new generation of software programs has high-quality feedback systems embedded within the program (Debuse & Lawley, 2016). A decade ago, Cassady and Gridley (2005) found that undergraduate students who took on-line practice tests performed better on their final exams. Formative assessments can serve the dual purpose of providing teachers with useful information about their students' knowledge and skills and familiarizing the students with the summative test environment.

The adoption of formative assessment presents another area for professional development in relation to implementing the CCSS. Schoenfeld (2015) noted that most

teachers have not been trained in the use of formative assessment. With professional development and training teachers can become adept in the use of formative assessment to prepare their students to meet the high expectations intrinsic to the CCSS as well as the challenges of the consortium-developed on-line summative assessments.

The *2012 MetLife Survey of the American Teacher* included questions about the CCSS. The survey involving 1,000 K-12 public school teachers and 500 principals was conducted in fall 2012 (Markow et al., 2013). Paradoxically, as preparing students for higher education has become more important fewer teachers believe their students are prepared for the college transition. While the vast majority (85%) of secondary school teachers felt that ensuring that every student graduates high school ready for college and a career should be a top priority, less than two-thirds (63%) believed that their students would graduate high school without needing remedial coursework in college.

The overwhelming majority of teachers (90%) and principals (92%) asserted that providing all students with core reading, writing, and mathematics skills would have a major impact on improving academic achievement and only slightly lower proportions (86% and 89% of teachers and principals, respectively) believed that high expectations would have a comparable impact (Markow et al., 2013). Markow et al. noted that at the time of the survey, the implementation of the CCSS in most schools and states was still in an early stage. At that time 46 states and the District of Columbia had implemented the standards. The survey sought to establish a benchmark for examining changes over time. More recent studies reveal that teachers were still experiencing challenges in implementing the standards (Burks et al., 2015; Hall et al., 2015).

Both teachers (92%) and principals (93%) considered themselves knowledgeable

or very knowledgeable about the CCSS although a higher proportion of teachers felt they were very knowledgeable (36% versus 20%). Markow et al. (2013) noted that the educators and administrators' confidence in their knowledge about the CCSS was similar across school levels and other sociodemographic attributes. While most teachers (59%) and principals (67%) viewed the implementation of the CCSS as challenging or very challenging, overwhelming majorities (90% of principals and 93% of teachers) were confident that the teachers in their schools possessed the academic abilities and skills to teach the CCSS. Although middle and high school principals were somewhat less likely to express that perspective, the teachers' beliefs were similar regardless of school level.

One area that teachers and principals both described as a challenge for school leaders was meeting the needs of diverse learners (Markow et al., 2013). This emerged as a particular concern for the teachers implementing the writing standards (Hall et al., 2015). Among the respondents to the *MetLife Survey,* educators in high-needs schools with high proportions of students performing below grade level expressed the greatest degree of concern (Markow et al., 2013). Close to two-thirds (62%) of the teachers reported that their schools were already using the CCSS to a substantial degree, especially at the elementary school level (68%). In addition, teachers in schools where most students were performing at or above grade level in mathematics and ELA were most likely to report that they were making more extensive use of the standards. However, principals of schools with high proportions of economically disadvantaged students were most likely to report that the teachers in their schools were using the standards extensively. It is possible that principals of low-income schools are more involved in the implementation of the CCSS.

Teachers using the CCSS in higher-achieving schools gave more attention to higher order skills in their lessons (Markow et al., 2013). Teachers in schools with more extensive implementation were most likely to report that teachers at their schools were devoting much of their instruction to teaching their students to read and comprehend literacy and informational material, apply mathematical knowledge to solving real world problems, reason abstractly and quantitatively, and to assess how point or view or purpose shapes the content and style of a text. According to the researchers it was evident how the school's current performance levels affected the challenge of successfully implementing the standards. It would seem that without additional resources, support, and assistance the schools might be perpetuating the achievement gaps that standards-based reforms were designed to address. NCLB did not eliminate achievement gaps and with the exception of elementary school mathematics showed negligible success in narrowing them (Dee & Jacob, 2010; Guisbond et al., 2012). The CCSS will not succeed in their goals of preparing all students for college and careers if teachers in high-performing schools in classrooms are primarily the ones to devote their instruction to higher order skills development and applications.

In fact, 80% of the teachers surveyed expressed the opinion that a variety of tools and supports would help them implement the CCSS with particular emphasis on teaching their students to solve real world problems (Markow et al., 2013). The resources and supports most teachers considered helpful included strategies and coaching to enable them to teach content more deeply, assessments aligned with the standards that signify mastery of content, tools to monitor individual student progress on key standards, simulations, games, and demonstrations of key concepts and principles, and more

effective techniques for homework and review. Notably, monitoring students' progress and games, simulations, and demonstrations are all techniques that are easily accessible with technology. Simulations and demonstrations are integral features of the on-line training teachers received to prepare them for administering on-line assessments (Rennie Center, 2015).

The teachers' responses revealed a discrepancy between their high levels of confidence in teachers' abilities to teach the CCSS and their lower confidence in the standards' potential impact on their students' educational outcomes (Markow et al., 2013). For the principals, their expectation that the standards would improve student outcomes was linked to their confidence in their teachers. That is, the more confidence the principals had in their teachers' ability to implement the standards, the more they believed that the standards would improve their students' academic achievement or make them better prepared for college and work. It is somewhat alarming that the middle and high school teachers, whose students were faced with mastering more complex knowledge and tasks and who were or would soon be in the process of preparing for higher education or a career were the least confident that the standards would improve their students' educational outcomes. According to Markow et al., the gaps between elementary and secondary school teachers' expectations and between the teachers' confidence in their abilities and the abilities of their students' have persistently surfaced in the teacher surveys.

Analogous to the association between principals' confidence in their teachers and their belief that the standards would improve educational outcomes, the teachers' perceptions of principal leadership influenced their use of and confidence in the standards

(Markow et al., 2013). Teachers who gave high ratings to their principals also made more extensive use of the standard and expressed more confidence in their potential benefits as well as their colleagues' ability to work with the standards. The principal sets the tone for the school. Principals who inspire confidence in their teachers are likely to establish a positive school culture with high expectations for the performance of all constituents.

Common Core State Standards and Consortia-Developed Assessments

According to Conley (2015), assessments can be depicted as falling along a continuum ranging from those that gauge "bits and pieces" of students' content knowledge to those designed to "capture student understanding in more integrated and holistic ways" (p. 14). Conley is highly critical of traditional educational assessments for failing to go beyond isolated "bits and pieces" of information. At the same time, he recognizes that multiple-choice tests have certain advantages, in particular, reliability and time-cost-efficiency. As a result, he does not envision an immediate end to multiple-choice testing, instead pointing out that the Common Core assessments combine multiple choice items with open-ended items, the format adopted for the Georgia Milestones Assessment System (GADOE, 2017).

The assessments aligned with the standards have been developed by two consortia, the PARCC and the Smarter Balanced Assessment Consortium (Conley, 2015). The assessments cover a broad range of Common Core standards for students in grades three to eight, with additional measures including performance tasks that capture more sophisticated understanding for students in high school. While agreeing that the consortia-developed assessments are superior to traditional standardized tests, Conley

argues that the new assessments still fail to capture specific competencies that students are expected to master by the time they graduate high school.

Conley (2015) recognizes that tests being developed by the PARCC and Smarter Balanced consortia have the potential to incorporate more sophisticated performance tasks than the ones being used. In fact, computerized tests can embed multimedia graphics, video, and audio clips that enrich the quality of the test material for engagement as well as the type of tasks on which students can be assessed (Schauffhauser, 2011). A key issue, however, is that some students will still be taking paper-and-pencil tests and for fairness and accuracy, the two modes of testing have to be equivalent. When switching from paper-and-pencil tests psychometric tests are administered to ensure the validity and reliability of the computerized test items and the equivalence of the two types of tests (Gewirtz, 2013; Hosseini, Abidin, & Baghdarnia, 2014; Kim & Huynh, 2007; Poggio, Glasnapp, & Yang, 2005). Therefore, until the states and school districts have completed the transition to on-line testing the assessments will not be able to exploit the full potential of the technology to provide more authentic assessments.

Nevertheless, the on-line assessments do make use of some features of the technology that extend their capacity for assessing performance. The Smarter Balanced tests use computer-adaptive testing which is designed to increase or decrease the degree of difficulty of the next question based on the test-taker's response to the previous question (Fink, 2015). This has the advantage of making the test more efficient as students are not wasting time answering questions that are too easy or too difficult for them. The test continues until the students have answered enough questions to be considered an accurate measurement of performance. The PARCC assessments utilize a

fixed format in which all students within a grade level are presented with precisely the same questions and tasks.

Colwell (2013) raised the question of whether test anxiety might affect the performance students taking computer-adaptive tests. While most questions center on the equivalence of the computerized and traditional paper exams, Colwell is skeptical of whether distinctions between the PARCC and the Smarter Balanced tests might give some students an unfair advantage. An earlier question was whether taking exams on-line would provoke test anxiety compared to taking the conventional paper test (Cassady & Gridley, 2005). However, findings from a comparison study found no evidence of that effect. Results from the first year of nationwide on-line testing revealed that students who took the PARCC test on paper tended to outscore those who took the on-line assessment (Herold, 2016). No similar discrepancy was reported for students who took the PARCC or Smarter Balanced exams, although the possibility might exist.

Both the PARCC and the Smarter Balanced tests include speaking and listening assessments although for districts using the PARCC, these features will be optional for the first few years (Fink, 2015). Students taking the tests with the speaking and listening components, which augment the writing and reading activities on the ELA assessment, are equipped with headsets which enable them to listen and respond to questions.

CAP outlined several challenges for the test developers and policymakers that are important for assuring that the standards and related assessment produce the intended improvements in educational outcomes (Polikoff, 2014). Most important, they assert that "Test developers in the consortia must put assessment quality and alignment issues front and center" (p. 3). The tests must be designed to "capture the full domain of the

standards, maintain the cognitive demand level of the standards content, and include a variety of high-quality items."

The concerns about the assessments are political as well as technical (Polikoff, 2014). As test developers and policymakers are promoting new uses of for assessment data, they "must provide reliability and validity evidence that the supports their intended uses to ensure that the appropriate decisions are made based on the assessment data" (p. 3). Although not mentioned directly, this implicitly refers to the use of test data for the evaluation of teachers and principals. The concerns raised by the Georgia teachers about the state teacher evaluation system (Owens, 2015) are not limited to any state. Teacher evaluation has persistently been a controversial topic (Darling-Hammond et al., 2012). By definition, high stakes assessments have a powerful impact on all stakeholders involved: "the progress and future of individual students; the reputations, salaries, and careers of teachers; and the reputations and status of schools as critical institutions within local communities" (Duffy et al., 2008, p. 53).

According to Conley (2015), the implementation of the CCSS and the aligned assessments, which has caused a number of states to reevaluate their achievement tests, may present an ideal opportunity for the states to advance to better models of assessment. He envisions a new "system of assessments" that synthesize multiple measures that go beyond meeting the states' accountability standards to provide students, educators, schools, and institutions of higher learning with valid information that *empowers* them to make prudent educational decisions (p. 23). A system of assessments involves *all* constituents who are affected by and who utilize assessment data. This invokes the concerns of CAP that the test data is used appropriately (Polikoff, 2014). However, both

sources believe that the new Common Core assessments are potentially far superior to previous generations of high-stakes assessments.

Review panels composed of practitioners, content experts, and specialists in assessment analyzed and evaluated the quality of four standardized assessments, including those developed by the PARCC and Smarter Balanced consortia, to determine whether they met the criteria delineated by the Council of Chief State School Officers (CCSO) for test quality (Doorey, N., & Polikoff, M., 2016). The two additional assessments are ACT Aspire and the Massachusetts Comprehensive Assessment System (MCAS). However, the PARCC and Smarter Balanced assessments are most relevant to this study. The analysis was driven by three key research questions: whether the assessments placed strong emphasis on key content for college and career readiness as called for by the CCSS and other standards (*Content)*; whether the assessments require all students to demonstrate the range of thinking skills, including higher order processes, called for by the standards (*Depth),* and what are the overall strengths and weaknesses of each assessment relative to the examined criteria for ELA and mathematics (*Overall Strengths and Weaknesses*).

Both Smarter Balanced and PARCC assessments were awarded an Excellent or Good Match rating on the CCSO criteria for the CCSS ELA and mathematics standards (Doorey & Polikoff, 2016). In addition, the panel agreed that the new assessments were far superior to previous state tests in emphasizing higher order skills. This was especially evident in the quality of the mathematics assessments. However, the highest rating (Excellent) was earned by the two assessments for ELA. Both mathematics assessments were rated Good Match.

In terms of Strengths and Weaknesses, the PARCC ELA assessment earned an Excellent Match on the CCCSO criteria related to the question of whether students are on track to meet college and career readiness standards (Doorey & Polikoff, 2016). According to the review panel the tests contain appropriately complex tests, encompass a range of cognitive load demand, and show sufficient variety in items types. One of the criticisms of the PARCC ELA is that it could be improved by including more research tasks requiring students to utilize multiple sources, and ultimately developing the capacity to assess speaking and listening skills. The PARCC includes this last feature, although it is currently optional (Fink, 2015). Conley (2015) argues that the assessments should involve more research tasks. For Content and Depth, the PARCC ELA assessment also earned the rating Excellent Match (Doorey & Polikoff, 2016).

For the mathematics assessment, the PARCC was rated Good Match on all three measures: alignment with college and career readiness standards, Content, and Depth (Doorey & Polikoff, 2016). The weaknesses of the PARCC mathematics assessment related primarily to the grade eight assessments. As to Content, the panel felt that the grade five assessments did not fully capture the major content at that grade level. Once again, weaknesses in the grade eight assessments prevented the test from receiving an Excellent Match on Depth. According to the panel, the grade eight tests were somewhat unbalanced, with a disproportionate number of items concentrated at higher levels of cognitive demand. The panel also critiqued the mathematics assessment on Depth for a number of problems related to editorial accuracy and to lesser extent, technical quality. This criticism highlights the high standards for quality set by the CCSO.

The Smarter Balanced ELA assessment was given a Good to Excellent Match in

relation to the college and career readiness standards (Doorey & Polikoff, 2016). The test was praised for capturing the most important aspects of the ELA skills specified by the CCSS, usually technology in ways that parallel real world applications, providing high quality measuring of the targeted skills. Notably, the program was deemed most successful in its assessment of writing and research and inquiry: three elements that unquestionably go beyond previous generations of tests. The assessment was also lauded for its unique inclusion of items that require active listening. On the other hand, the review team felt that vocabulary items warranted improvement, as did the cognitive demand in grade five items. They also recommended that over time, the test could develop the capacity to assess speaking skills. On Content and Depth, the Smarter Balanced ELA program earned ratings of Excellent Match and Good Match, respectively.

The Smarter Balanced mathematics assessment was rated Good Match in all three areas (Doorey & Polikoff, 2016). Criticism focused primarily on the grade five and grade eight programs. On Depth the criticisms were analogous to those expressed for the PARCC mathematics program. Overall, the flaws and weaknesses discovered by the review team were relatively minor and could easily be addressed. The overarching implication of the analysis was these tests represent a dramatic improvement over previous standardized assessments that barely (if at all) tap into higher order skills. For the most part, the tests are aligned with the CCSS and meet the criteria demarcated by the CCSO. With the tests deemed suitable if not excellent, the challenge facing the education community is the district level implementation of the online assessments.

District level implementation

In spring 2014, the Center for Education Policy (CEP) conducted a survey of public school districts to investigate their progress in implementing the new assessments and the challenges they confronted. The report presented by Rentner, D.D., Kober, N., Ferguson, M., McMurrer, J., and Frizzell, M., (2014) was based on data drawn from a subset of districts involved in the larger survey. At the time of the survey the districts were still uncertain about the impact of the assessments on improving teaching and learning and providing important information for stakeholders. Close to half the districts felt it was still "too soon to tell" whether the assessments would produce information to guide instruction in mathematics or ELA (48% and 46%, respectively). Slightly more than half (54%) of the districts felt it was too soon to determine whether the assessments would constitute an improvement over the state's existing assessments, but they were optimistic; 55% thought the new assessments would drive instruction in positive ways and 55% believed they would yield results that could be understood and used by parents and students (64%).

Schoenfeld (2015) extolled the benefits of formative assessment. As a result of the new assessments more than half the districts (53%) were contemplating reviving the formative assessments in ELA and mathematics that teachers used to guide their instructional practice (Rentner et al., 2014). In addition, roughly 45% of the districts were considering whether to revise their interim assessments which are used to gauge whether students are on track to perform well on future high stakes assessments. Very few districts considered eliminating their current formative or interim assessments.

Technology issues emerged as the major challenge for the districts. Roughly

three-quarters (76%) of the districts reported experiencing major or minor problems with enough computers with adequate processing speed and other features to administer the CCSS assessments to large numbers of students (Rentner et al., 2014). Similar proportions of districts reported major or minor challenges with finding sufficient numbers of staff at the district or school levels possessing the expertise needed to address technology-related problems that might surface while the tests were being administered. This issue emerged as a common complaint among teachers involved in field testing the PARCC assessment (Rennie Center, 2015). More than 60% of the educators who acted as test administrators felt their training poorly prepared them to deal with even basic technology issues. As a result of focus group discussions both sites where the field tests were held concluded that a cadre of teachers should be trained to serve as "tech troubleshooters" at future test sessions. The teachers would serve as frontline technology support, boosting the support provided by tech support at the school level and freeing district technology support staff.

A more serious problem, less easily addressed than providing teachers with technology training or seeking staff members with technical expertise, was the absence of the technology infrastructure needed to support large-scale on-line testing (Rentner et al., 2015). More than half the districts (55%) did not expect to have the requisite infrastructure until the 2014-2015 academic year or later. A majority of districts expressed concerns about adequate bandwidth and internet access in schools, with about 29% describing it as a major challenge. The issue of having security measures in place to protect the on-line assessments from cheating was deemed a major challenge by 14% of the districts and a minor challenge by 38%. Overall, only 32% of the districts reported

having the necessary infrastructure for on-line testing at the time of the survey.

According to Mel Riddle, a principal and Associate Director of High School Services for NASSP, other surveys have also found that most districts were not prepared for massive on-line test administration (Riddle, 2015b). Most schools had to close their libraries and computer labs to devote all available technology to testing, often lasting a full month. Inadequate infrastructure is a particular problem in rural and low-income school districts. Riddle's school conducted speed tests and stress tests on all computers before the testing season began. They learned from experience to proactively troubleshoot potential technology problems.

The overwhelming majority of the districts surveyed had or were in the process of making plans to provide support services for students who might require additional support to pass the Common Core assessments (Rentner et al., 2015). These students included students who had achieved the proficient levels on the state ELA or mathematics assessments but were still likely to need assistance passing the more rigorous exams as well as students who performed below the proficient level on prior math or ELA assessments. The challenge for many districts (25%) was that they lacked adequate funding to carry out their plans for supporting the at-risk learners.

The survey provided the district officials with space to write in additional comments about the prospective impact and utility of the new assessments (Rentner et al., 2015). One concern was that the assessments might be more of an assessment of the students' technology skills than their knowledge of content. Some educators view young learners as "digital natives" who would have no problem adapting to the computerized tests. In reality, students who are adept with using computers for homework assignments,

video games, social media, and other personal uses are not necessarily prepared for the rigors of on-line assessment (Gewirtz, 2013, 2015; Riddle, 2015a, 2015b; Schauffhauser, 2013). Riddle (2015b) observed that test scores dropped the first time students were given exams on-line. He advocates providing students with repeated practice opportunities to prepare them with on-line testing.

Some educators are especially concerned with whether younger children have the skills to participate in on-line testing (Schauffhauser, 2013). For example, they suggest that young learners might have difficulty highlighting and moving text. However, test developers assert that the systems are designed to not require a great deal of dexterity. To ensure that the tests are user-friendly for children AIR (American Institute of Research) regularly conducts cognitive labs, typically with third or fourth graders. According to one representative from AIR, the children proved to be highly adept. District officials are more skeptical. A pilot study with Smarter Balanced convinced one district to play closer attention to the students' technology skills and ensure that they have the requisite skills for performing the basic computer tasks for on-line testing.

In the months moving up to the first on-line assessments, many schools began providing their students with intensive instruction in basic computer skills such as keyboarding, scrolling, using the cursor, and dragging and dropping text (Gewirtz, 2015). Ironically, after complaining that NCLB was turning classrooms into "test prep" centers, teachers were devoting instructional time to "tech prep." According to interviews conducted by *Education Week,* some teachers viewed the "tech prep" as a waste of time, but they were outnumbered by the teachers who considered teaching students computer skills a worthwhile investment, aligned with the CCSS objectives to prepare students for

college and careers in a technology-driven world. Furthermore, many teachers argued that computer skills should be integrated into subject content instruction rather than taught in isolation. Whereas the AIR representative interviewed by Schauffhauser (2013) said they were "amazed" at the young students' expertise with computers, the teachers were more likely to report being amazed by their students' *lack* of technology skills (Gewirtz, 2015).

As Cassady and Gridley (2007) pointed out a decade ago, on-line formative assessments simultaneously familiarize students with the testing format while providing students and teachers with the benefits of formative assessment per se. Time emerged as a major concern for many teachers who were trying to fit computer skills training into the school day without detracting from regular lessons (Gewirtz, 2015). Integrating computer skills and course content seems the most practical and effective way to ensure that students are prepared for all facets of on-line high-stakes testing.

A persistent concern about assessments aligned with the CCSS is that the more stringent standards would result in declines in test scores. In fact, that was the case in many school districts, causing educators to question the merits of the new tests (Riddle, 2015a). Preliminary results from the Georgia Milestones revealed that fewer students would classify as proficient (GADOE, 2017). However, the GADOE points out that on other indicators such as the Georgia NAEP, the SAT, and the ACT, the state's students have demonstrated improvement.

Riddle's (2015a) response to educators who wanted to return to prior assessments was to inform them that only 3% to 10% of the traditional tests measured higher order skills. In contrast, analyses of the PARCC and Smarter Balanced assessments showed

that 61% of the mathematics assessments and 78% of the ELA tests captured higher order skills. To Riddle, the conventional multiple choice tests might be a validation of a high school diploma, but they do not signify college and career readiness. The main selling points of multiple choice tests are their accuracy, cost-efficiency, and ease of use (Clarke et al., 2000).

After the initial dip in test scores on the Georgia Milestones, the GADOE (2016) reported that students' performance in most subjects increased in most subject areas. The Georgia Milestones differs from the previous state assessment in that it outlines four levels of achievement rather than three. *Beginning learners* are students who have not demonstrated proficiency in the knowledge and skills stipulated by Georgia's content standards. Students who score at this level require substantial academic support to progress. *Developing learners* demonstrate partial proficiency in the requisite knowledge and skills and require additional but less intensive support to advance. *Proficient learners* demonstrate the requisite level of knowledge and skills and are considered sufficiently prepared for the next grade or course and on track for college and career readiness. *Distinguished learners* demonstrate advance proficiency and are well prepared to advance. Students must meet more stringent criteria for proficiency, which at least partly accounts for the smaller proportion of students who reached proficiency the first year. On the 2016 assessment the proportion of students achieving the status of proficient learner or above increased for 23 of the 32 tests.

Navigating the Transition to On-line Testing

A report by EdTech (2015) cited several reasons for making the shift from paper-and-pencil to on-line testing. Advantages offered by on-line testing include: (a) enhanced

opportunities for effectively gauging students' understanding and performance; (b) enhanced accessibility for students with special needs; (c) more efficient test delivery and much quicker turnaround of results; (d) an improved security model; and (e) increased student motivation and engagement. Despite these potential advantages, the report acknowledged that seven states experienced significant-technology related testing failures resulting in fines being levied against four test vendors. In each case the vendors were delivering a new online assessment or an assessment with new test functionality. Given this occurrence, additional states with contracts for new on-line tests may be at heightened risk for similar malfunctions.

The testing failures resulting from the use of specific assessments for the first time were not the first or only technology-related incidents to cause trouble for schools during testing. The previous year Florida, Kansas, and Oklahoma were forced to suspend on-line testing due to malfunctions that dramatically slowed down the system or locked students out of the program (Davis, 2014). It was the second year in a row that states experienced serious disruptions to on-line testing procedures. The negative publicity generated enough anxiety that some districts decided to return to pen-and-paper assessments.

By 2016, after on-line testing had gone nationwide, the number of states reporting serious technology-related problems, in some cases forcing them to abandon the on-line testing system and start from scratch, continued to increase, creating more stress in the education community. Problems in Florida and Kentucky were eventually connected to cyber attacks (Doran, 2016).

Alaska, Indiana, Minnesota, and Tennessee all experienced server crashes or other

malfunctions (Burnette, 2016). The most publicized case was in Tennessee (Camera, 2016a, 2016b). On the morning of February 8, 2016, as districts across the state were beginning their test administration it soon became apparent that the whole system was crashing. One reason for the publicity was that the state and some of its districts were leaders in the Race to the Top. Tennessee was in the vanguard in the adoption of the CCSS and the Kingsport district was the first in the state to fully implement the standards, a year before it was required. The fact that the technology breakdown occurred in a state that was leading the nation in the new assessment system elevated apprehension about switching to on-line assessments. Tennessee briefly returned to paper-and-pencil assessments and subsequently rebooted the on-line assessments (Doran, 2016).

In addition to the lawsuits brought against vendors, teachers unions and anti-testing advocates in affected states have inundated legislators with letters from students and teachers vividly describing the "emotional turmoil" caused by last-minute technology breakdowns (Burnette, 2016, p. 16). Not unexpectedly, the testing companies and their allies immediately countered that the unions and other groups were simply capitalizing on the malfunctions as part of a drive to eliminate standardized testing.

Adding to the controversy was the recent passage of ESSA endowing the states with greater flexibility in their approaches to testing (Burnette, 2016). State legislators in Indiana and Minnesota were considering discarding the state assessment system and starting over. As the states began the transition to on-line testing, legal experts cautioned that legal issues would arise (Bathon, 2013). The problems that have occurred have spurred states and districts to proactively troubleshoot their technology systems and

bolster their security. The assessment programs are still evolving and the systems in many states are still in a nascent state.

Paper-and-Pencil Testing vs. On-line Testing

One of the biggest concerns in shifting to on-line assessments from paper-and-pencil tests is that the two assessments are equivalent, especially as some students are still being tested in the traditional format (Gewirtz, 2013). The states have traditionally relied on statistical and psychometric analyzes for updating paper-and-pencil tests. However, the present transition to the consortia-developed tests is more challenging. The addition of interactive and graphic items and performs tasks on the computerized tests heightens the complexity.

Fears that the two test formats might not be equivalent and therefore might favor one group seemed to be realized with reports that students who took the 2014-2015 PARCC on-line tended to score lower than those who took the paper-and-pencil exam (Herold, 2016b). Apparent differences in performance between the two tests varied across districts and in some case were negligible or nonexistent. In other states and districts the differences were substantial, as in Illinois where the state board discovered that 43% of the students who took the ELA on paper scored as proficient or above compared to 36% who took the ELA on-line. A similar pattern emerged in Maryland, where middle grade students in Baltimore scored nearly 14 points higher on the PARCC ELA compared to peers with equivalent academic and sociodemographic profiles who took the ELA on-line. This evidence of a "mode effect" questioned the validity of the test.

The National Center for Educational Statistics (NCES) acknowledged that they

found some evidence of a subgroup effect in transforming the NAEP from a paper to a computerized test (Herold, 2016b). Their investigation suggested that the students' prior experience with computers was the cause of the discrepancy. In Baltimore County, where the differences were pronounced, district officials performed further analyses to identify the reason for the discrepancies by comparing how students and schools with similar academic and background characteristics performed on each version of the test. They discovered that after controlling for student and school characteristics, the students were between 3% and 9% more likely to score as proficient on the paper-and-pencil mathematics test with the magnitude varying by grade level. Students taking the paper-and-pencil ELA exam were 11% to 14% more likely to score proficient. The Baltimore County accountability chief commented that the disparities would make it difficult to draw comparisons for the first and second years of the test. Nevertheless, PARCC insisted that the test still provided valuable information for informing instructional improvements and identifying students in need of assistance or enrichment.

The discrepancies in the test scores and the technological problems that plagued a number of states added to an atmosphere of uncertainty surrounding the transition from paper to on-line assessments. However, several states have been performing on-line assessments for several years. Hawaii switched to on-line testing for 2010-2011 after piloting on-line assessments a year earlier, only to discover that the school district (the entire state) did not have sufficient bandwidth or computers (Schauffhauser, 2013). Convinced of the merits of on-line assessments, technology leaders decided the problems were not insurmountable. Other states encountered obstacles in the beginning, but like Hawaii, they worked to resolve their problems. Given that the consortia-developed tests

are aligned with the goals of the CCSS to promote the development of higher order skills and prepare students for success in real world endeavors, the on-line assessments have far greater capability for measuring students' performance on problem solving and authentic tasks. In the cost-benefit analysis the advantages of the on-line assessments outweigh the drawbacks which are likely to be temporary.

In addition to the question of whether paper and on-line assessments produce equivalent results is the question of prospective differences in performance on different devices. A report commissioned by the CCSS provided a comprehensive analysis of the numerous features of technological devices that could potentially affect test performance (DePascale, Dadey, & Lyons, 2016). The authors note that the literature on score comparability is "broad and varied" and the definitions of score comparability are likewise (p. 5). For their purposes, DePascale et al. turned to Winter's definition which is based on the premise that comparability requires that a test and its different versions must: (a) measure the same set of knowledge and skills at the same level of content-related complexity (constructs); (2) produce scores at a stipulated level of specificity that captures the same degree of achievement on those constructs; and (3) have similar properties (reliability, decision consistency, sub score relationships) relative to the level of score reported. According to DePascale et al., meeting those criteria is a challenging prospect but various sources of evidence can be used to support score compatibility.

With respect to the comparability of testing on different devices there are numerous features of the technology that can impact performance (DePascale et al., 2016). These include screen size, input mechanism (touch screen or mouse), and keyboard (onscreen or external). Device positioning and ergonomics can also affect

performance. In experiments where children took tests on tablets, Laughlin Davis, L., Strain-Seymour, E., and Gay, H., (2013) observed that the test-takers either laid the tablet flat on the desk, or with an iPad, propped it up on its cover. In contrast, adults in similar studies took full advantage of their device's display options and adjusted their bodily positions for greater efficiency and comfort.

As the comparisons between adults and children illustrate, the characteristics of the user interact with characteristics of the device. In fact, DePascale et al. (2016) turned to the work of Laughlin Davis et al. (2013) in noting that comfort and familiarity are user characteristics often found in research. The NCES (National Center for Educational Statistics) ascribed differences in NAEP (National Assessment of Educational Progress) scores on paper-and-pencil and on-line tests to familiarity with computers (Herold, 2016b). Sponsored by Pearson, the studies presented by Laughlin Davis et al. (2013) and Strain-Seymour, E., Craft, J., Laughlin Davis, L. and Elbom, J., (2013) were designed to investigate the utility of tablets in K-12 assessments. The researchers examine the impact of the device features outlined by DePascale et al. (2016).

Strain-Seymour et al. (2013) described their experimental study as a hybrid between a usability study and a cognitive laboratory. The participants were 24 students from two Virginia school districts selected to represent three grade levels: grade four, grade eight, and high school. All the students were familiar with the Virginia Standards of Learning (SOL) and with on-line testing but only one of the districts used tablets and mobile devices (iPads and iPods) in lessons. For the study, the students used 10" Samsung Galaxy tablets with Bluetooth external keyboards. Using a think-aloud protocol, the students were tested individually on select items drawn from the SOL

involving: multiple choice answers, “hot spot” items involving the selection of one or more elements or areas of an image, drag and drop items, passages displayed via a paging interface, tools (highlighters, pencil tool, answer eliminator), navigational controls, and an essay-writing interface.

Although the students had never taken a test on a touch screen before, their familiarity with the on-line testing interface facilitated their ability to transfer their navigational skills to the device (Strain-Seymour et al., 2013). The researchers observed highly enthusiastic reactions to using the tablet, adding a motivational element to taking a technology-based test (Chua & Don, 2013). As far as device features, the touch screen interfaces allowed direct and immediate input but with less precision than using a mouse and fewer channels for user feedback (Strain-Seymour et al., 2013). Using the touch screen also demanded greater visual attention than an external keyboard even for students with more keyboarding expertise. One drawback was related to text editing, which the students found difficult due to the small screen size.

An unexpected finding was the preference of younger students and novice typists for the onscreen keyboard and difficulties all students encountered with the external keyboard, including expert typists (Strain-Seymour et al., 2013). The students found the external keyboard awkward and frustrating and it was not fully compatible with the device. The study was part of ongoing research into the viability of using tablets for testing, which is an important issue in view of the inadequate number of computers in many school districts (Rentner et al., 2014). The overarching finding was that the students encountered technical issues in using the tablets for specific test elements but at the same time they were excited by the novelty of using the tablets for tests (Laughlin

Davis et al., 2013; Strain-Seymour et al., 2013). "Cool" and "more fun" were common reactions despite some obvious frustration with the technology. In fact, Laughlin Davis et al. (2013) noted that the students preferred tablet tests to either computer or pencil-and-paper tests even when they earned lower scores. The ongoing studies are aimed at optimizing the user interface and peripherals of tablets for test purposes. However, Laughlin Davis et al. cautioned against using tablets for high stakes testing without first conducting large-scale comparisons involving tablets, computers, and paper tests.

In the time since the Pearson lab studies were conducted, tablets have been used in high stakes testing. The results seem to support the observations of Laughlin Davis et al. (2013) that some test questions are more difficult to answer on tablets, potentially impacting students' scores (Herold, 2016a). Compounding the discrepancies in scores between students who took the PARCC test on-line or on paper (Herold, 2016b), evidence also emerged of a "device" effect for using tablets to take the PARCC assessment. For example, one analysis reported that students in Ohio (14% of all students who took the 2014-2015 PARCC) scored significantly lower taking the test on tablets (Herold, 2016b).

Reports issued by both PARCC and Smarter Balanced concluded that their tests produce comparable results regardless of mode of assessment (Herold, 2016b). However, detailed analyses of tasks and test questions on the PARCC indicate that a number of items on the Algebra 1 and geometry exams were flagged as being more difficult for students taking the test on tablets and some Algebra 2 items were flagged as more difficult to complete on a computer. Analyses of raw scores suggested evidence of a device effect, but most striking, the researchers observed Ohio students who took the

PARCC end of year and performance-based tests on tablets scored an average of 10 points lower than those who took the exams on laptops and 14 points lower than those who took their exams on desktops. Described as "atypical," these discrepancies seemed to be unique to Ohio and seemed to defy explanation.

Without any definitive explanation, PARCC officials surmised that the test results might have been skewed by a large district that relied heavily on tablets and where students were disproportionately underperforming or unfamiliar with the devices (Herold, 2016a). The CCSS report on comparability noted that several studies have disclosed evidence of device effects (DePascale et al., 2016). They called for ongoing analyses of test results and issued best practice guidelines for the states for the purpose of minimizing and monitoring threats to comparability.

Thus far there have only been two spring term assessments where the vast majority of United States students took technology-based exams. The discrepancy between results on the on-line and paper tests raised apprehension, but computerized tests and psychometric analyses of equivalence have a long history (Gewirtz, 2013; Hosseini et al., 2014; Kim & Huynh, 2007; Poggio et al., 2005). The use of tablets in testing is recent, and as the CCSSO and the Pearson researchers recommend, should be subject to ongoing, detailed, large-scale quantitative analysis (DePascale et al., 2016; Laughlin Davis et al., 2013; Strain-Seymour et al., 2013).

Studies examining students' responses to computerized testing have been conducted for decades, primarily involving undergraduate and graduate students. In a recent study, Boevé, A.J., Meijer, R. R., Albers, C. J., Beetsma,Y., and Bosker, R. J., (2015) presented a field test of computerized high stakes testing involving 400

psychology students. The students were randomly assigned to take the midterm either on paper or on the computer. For the final exam, the students who took the computer-based exam took the paper test and vice versa. A survey assessing acceptance of computer-based exams queried the students on their experiences, their preferred mode of testing, and whether their opinions of computerized exams had changed as a result of their experience.

No differences emerged in the students' test performances as a result of their taking the computer or paper test (Boevé et al., 2015). However, the students tended to prefer the traditional paper test, which seemed to allow them to feel more in control. Roughly half the students preferred the paper-and-pencil test and one-quarter expressed a preference for the computerized test. Nevertheless, 43% of the students felt more positive toward computer-based testing after having experienced it compared to only 14% who expressed a more negative opinion. The remaining students did not change their opinion. Boevé et al. attributed the most positive perceptions to the fact that the computer provided immediate feedback on test performance. This feature has long been considered an advantage of computerized testing (Cassady & Gridley, 2005).

Chua and Don (2013) explored motivational effects of computer-based testing as well as performance in a study of teacher education students taking a multiple choice biology test. Based on prior performance and computer skills, the students were divided into four equivalent groups, two taking computer-based tests and two taking paper-and-pencil tests. The study included a Biology Motivation Questionnaire (BMQ). Comparisons of the two modes of testing disclosed significantly lower performance on the paper-and-pencil test. Motivational effects might have accounted for this difference.

Students who took the computer-based tests displayed increased intrinsic motivation and self-efficacy although they also had higher anxiety. Younger students may be even more prone to motivational effects, which were clearly evident in the tablet studies (Laughlin Davis et al., 2013; Strain-Seymour et al., 2013).

Language tests were among the first to shift to computers. The Test of English as a Foreign Language (TOEFL) began transitioning to computer-adaptive testing in 1998 (Kim & Huynh, 2007). Hosseini et al. (2014) compared the reading comprehension performance of Iranian EFL undergraduate students on computer-based and paper-and-pencil tests. The participants were 106 first year students who were also surveyed on their attitudes toward and experiences with computers. Analyses revealed significantly higher scores on the paper-and-pencil tests and significantly more variations in performance on the computer-based tests. Despite the effect on performance, more than half the students (53%) preferred the computer-based tests. Those who expressed this preference found the computer-based test more comfortable, more enjoyable, and more time efficient. Features that facilitated ease of use made the technology-based tests more attractive to many students. Hosseini et al. (2014) and Chua and Don (2013) both noted that the differences they found between the two test formats contrast with most studies, which typically find no significant differences in performance on the two types of tests.

Georgia Case Study

Ogletree et al. (2014) presented the case study of the third author Allen, a principal of a high-needs urban elementary school involved in the transition from paper to on-line testing. In the innovative school model, the students are heavily involved in research-centered assignments within a structured interdisciplinary curriculum. To

prepare the students for on-line testing, the school administered three benchmark tests over a nine-month period. The tests allowed the staff to discover that the youngest learners had difficulty navigating the screen as well as other challenges that needed to be addressed.

Another essential step in preparing for on-line testing involved taking an inventory of the available technology, which included the media specialist, technology support specialist, testing coordinator, and principal working collaboratively (Ogletree et al., 2014). The team discovered that only a limited number of devices were available for testing, thus indicating that scheduling and implementation would be a complex process. This has been a common problem in school districts (Rentner, 2014). The number of devices allowed two classes to be tested at the same time, one in the computer lab and the other in their regular classroom (Ogletree et al., 2014). This raised issues related to the testing environment, as the team surmised that the students taking the test in the familiar classroom environment might feel more confident and at ease.

Security is a major issue in the transition to on-line testing, although Riddle (2015b) found it to be less of a problem than having to lock up reams of paper tests to prevent cheating. The school team devised a monitoring system with the same degree of security as customarily had for standardized tests (Ogletree et al., 2014). The nine-month preparation period enabled the school to meet challenges that emerged with logistics, hardware and software and allow students to become familiar with on-line testing. Through strategic planning the school was able to overcome hurdles to implementation that are common to schools and districts shifting from paper to on-line tests (Gewirtz,2013; Rentner et al., 2014; Schauffhauser, 2011, 2013).

Chapter Summary

Descriptions of NCLB range from well-intentioned but misguided to discriminatory and detrimental, but the overall conclusion is that it was unsuccessful in improving student achieving and closing achievement gaps. The CCSS represents a dramatic departure from prior standards in that it emphasizes higher order skills and real world applications of knowledge and skills. In an additional shift from traditional practices, for the 2014-2015 school year, districts in all states transitioned from paper-and-pencil to on-line assessments for the annual state-mandated tests. The assessments were developed by two consortia, PARCC and Smarter Balanced (Conley, 2015). Many districts felt unprepared for the shift, especially with regard to adequate infrastructure and computers (Rentner et al., 2014). Some apprehensions were realized when several states experienced severe disruptions, in some cases forcing them to suspend on-line testing (Camera, 2016a, 2016b). Furthermore, results in some districts showed higher scores for students who took the paper-and-pencil PARCC assessment (Herold, 2016a, 2016b). The CCSSO, which is largely responsible for the development of the CCSS, recommends ongoing analysis to ensure that the online tests are fair and equivalent to the traditional paper tests (DePascale et al., 2016).

In response to unduly high rates of attrition among teachers in Georgia, the GADOE conducted a survey to discover the sources of teachers' dissatisfaction (Owens, 2015). High stakes testing emerged as the foremost source of dissatisfaction, followed by the connection of test results to teacher evaluation. To ensure that our schools have excellent teachers, it is imperative to gain insight into teachers' and principals' perceptions of changes affecting their professional practice. This study will explore the

perceptions of teachers and principals in a rural Georgia school district toward the Georgia Milestones Assessment System and the transition to on-line testing.

CHAPTER III

METHODOLOGY

No Child Left Behind (NCLB) was a standards-based education reform that required states to institute high stakes, standardized testing each and every year in grades 3-8 in order to gauge the level of student knowledge and achievement (Smith, J. M. & Kovacs, P. E., 2011). However, NCLB fell short of realizing all of its goals. Recognizing this fact, President Barack Obama signed the Every Student Succeeds Act (ESSA) into law.

This reform effort brought about the implementation of a new standardized testing system, the Georgia Milestones Assessment System (GMAS). Because the GMAS is administered primarily on-line, it brought about new concerns holding schools and teachers accountable for providing high quality education to all students. The GMAS reform effort is contingent upon what happens when the GMAS as an on-line high stakes test is fully implemented.

Research suggests that how teachers and principals implement a policy is influenced by their perceptions of policy (Honig, 2006). Although there is limited research regarding on-line high stakes testing in the elementary grades, there is noted absence of research regarding teachers' and principals' perception regarding on-line testing of the Georgia Milestones Assessment. This chapter discusses the research methods and procedures used to research grades 3 – 5 teachers' and principals' perceptions regarding Georgia Milestones on-line testing in a Georgia school district.

Research Question

This research addressed the following research question: How do grades 3-5

principals and teachers perceive Georgia Milestones on-line testing in a Georgia school district?

Research Design

This study utilized quantitative research methods. In quantitative research the goal is to establish the relationship between one variable (an independent variable) and another (dependent variable) within a population (Hopkins, 2008). The independent variable is the respondents (teachers and principals). The dependent variable is the answer to each question on the survey. Data collection via survey was chosen so that the data could be collected in an efficient and cost effective manner.

Participants

This study was conducted in a rural Georgia school district serving approximately 900 elementary (K-5) students. The population of the school district is approximately 30,000 with a median household income of $40,000 (U.S Census Bureau). Two grades 3 – 5 elementary schools are located in the school district. School A houses 15 certified teachers and School B houses 45 certified teachers. All 60 teachers completed surveys. School principals at each school who were asked to participate were Principal, Assistant Principal, Assistant Principal of Instruction, and an Assistant Principal of Discipline. All 8 principals completed the survey.

Instrumentation

The participants completed a survey based on a modified version of the *Teacher's Beliefs about the Effects of High Stakes Testing Survey* developed by Brockmeier, Pate, and Leech (2008). Brockmeier et al. survey consisted of forty-nine Likert questions. The range was from (1) (strongly agree) to (5) (strongly disagree). The survey assessed forty-

nine items comprised of six hypothesized subscales (i.e. curriculum, teaching, work satisfaction, stress, accountability, and students). Subsequently, Brockmeier, Pate, and Leech (2008), provided and in-depth analysis of the psychometric characteristics (i,e, validity and reliability) of the survey. For this study, only the curriculum, teaching, stress, and accountability domains were used. No questions were reverse-scored. Four open-ended questions were added at the end of the survey to collect additional data. The only modification made to the survey was 'high stakes testing' was changed to 'on-line high stakes testing'. Permission to use the instrument was obtained via email (Appendix C).

Data Collection

The researcher obtained Institutional Review Board (IRB) of Georgia Southern University and school district permission. Upon approval, the researcher communicated with the principal of each school requesting to be placed on the agenda for an upcoming faculty meeting and a mutually agreeable date was selected. At the faculty meeting, the surveys were distributed and directions for completing the survey were explained by the researcher. An envelope was given to each teacher. Each envelope contained an informed consent letter and a survey. The researcher returned to the office area and waited for a designee of the principal to collect the survey. The principal designee placed all surveys in a large envelope and gave them to the researcher. Survey and open-ended questions were then analyzed.

Data Analysis

Data were analyzed using SPSS. Descriptive analysis (mean, standard deviation) by question was determined for each group, teachers, and principals. Descriptive analysis (mean, standard deviation) by domain was determined for each group principals

and teachers. Data trends for each group by question and domain was determined. Open-ended questions were analyzed for common domains/themes for each question. Once the data had been collected and analyzed for the study, it was placed in a locked file cabinet to be kept for five years. After that time, the data will be shredded and burned.

Chapter Summary

The researcher used the data collected to describe characteristics of the perceptions as well as the differences in these perceptions of Grades 3-5 teachers and principals regarding on-line testing of the Georgia Milestones high stakes test in a rural Georgia school district. The methodology was descriptive utilizing quantitative methods which encompassed administering a survey to teachers and principals in two schools in one public rural school system. Permission was obtained from the Institutional Review Board of Georgia Southern University before the survey was conducted. The study used a modified version of the Teacher's High Stakes Testing Survey that included eighteen questions and 4 open ended questions as the main method of collecting data that was used in developing conclusions. The survey was given to the participants and returned anonymously. Participation was voluntary. Data was analyzed consisting of descriptive statistics to establish associations between variables.

CHAPTER IV

REPORT OF DATA AND DATA ANALYSIS

A quantitative study was conducted to determine grades 3 – 5 teachers' and principals' perceptions regarding Georgia Milestones on-line testing in a Georgia school district. Quantitative data were collected from two elementary schools and eight administrators (4 in each school). Data were collected via Likert scale survey items and open-ended questions at the end of the survey. This chapter presents the findings of the research study.

Findings

This section provides an analysis of the research question. Descriptive statistics (mean and standard deviation) were completed for each survey domain by teachers and principals. Domains were: (1) Curriculum, (2) Teaching, (3) Stress, and (4) Accountability. Findings are presented by domain. Open-ended questions were analyzed using a common theme approach by teacher and principal. Findings are presented by teacher and principal.

How do grades 3-5 teachers and principals perceive Georgia Milestones on-line testing in a Georgia school district?

The researcher calculated descriptive statistics (mean and standard deviation) for survey question that aligned to each domain. Teacher data were analyzed by school and combined data. Principal data were analyzed by school and combined data. Data are presented by domain with an overall findings discussion at the end of domain analyses. Open-ended survey questions were analyzed for common domains/themes. Data are presented by question by teacher and principal.

Domain 1: Curriculum

The first five questions in the survey mapped to the curriculum domain in both the teacher and principal surveys. The curriculum domain captured teacher's thoughts about the relationship between the Georgia Milestones Assessment and the state curriculum (Table 1).

Table 1

Mean and Standard Deviation for Questions in the Curriculum Domain by Teacher

Item #	Question	School A (N = 45)		School B (N = 15)		Combined (N = 60)	
		M	SD	M	SD	M	SD
1	On-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach.	1.22	0.42	1.06	0.25	1.18	0.39
2	On-line high stakes testing is counter to the idea of a balanced curriculum (equal attention to subjects).	1.11	0.32	1.33	0.48	1.16	0.37
3	On-line high stakes test items accurately reflect the content students learn through a school's curriculum.	3.11	0.32	2.33	0.48	3	0.71
4	Students' scores on on-line high stakes tests provide feedback for schools to improve the curriculum.	3.68	0.8	3.6	0.73	3.66	0.7
5	On-line high stakes test content is aligned with the school's curriculum.	1.84	0.36	1.8	0.41	1.83	0.37

School A analyses found teachers agree to strongly agree on-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach (M = 1.22), does not lead to a balanced curriculum approach (M = 1.11), and testing content is aligned with their school's curriculum (M = 1.84). Teachers do not agree on-line high stakes test items accurately reflect the content students learn through a school's curriculum (M = 3.11) or students' scores on on-line high stakes tests provide feedback for schools to improve the curriculum (M = 3.68). School B teachers are in

agreement with School A faculty on 4 of the 5 survey questions (1, 2, 4, and 5).

Table 2

Mean and Standard Deviation for Questions in the Curriculum Domain by Principal

Item #	Question	School A (N = 4)		School B (N = 4)		Combined (N = 8)	
		M	SD	M	SD	M	SD
1	On-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach.	2.0	0.0	1.5	0.57	1.75	0.46
2	On-line high stakes testing is counter to the idea of a balanced curriculum (equal attention to subjects).	1.5	0.57	2.0	0.0	1.75	0.46
3	On-line high stakes test items accurately reflect the content students learn through a school's curriculum.	1.0	0.0	1.0	0.0	1.0	0.0
4	Students' scores on on-line high stakes tests provide feedback for schools to improve the curriculum.	1.0	0.0	1.0	0.0	1.0	0.0
5	On-line high stakes test content is aligned with the school's curriculum.	4.25	0.57	4.0	0.0	4.13	0.35

School A analyses found principals agree to strongly agree on-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach (M = 2.0), does not lead to a balanced curriculum approach (M = 1.5), accurately reflect the content students learn through the school's curriculum (M = 1.0), and that on-line high stakes tests provide feedback for schools to improve the curriculum (M = 1.0). Principals do not agree on-line high stakes test content is aligned with the school's curriculum (M = 4.25). School B principals are in agreement with School A faculty on 5 of the 5 survey questions (1, 2, 3, 4, and 5).

Overall, teachers and principals agree that on-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach (teacher

mean = 1.18, principal mean = 1.75) and does not lead to a balanced curriculum approach teacher (mean = 1.16, principal mean = 1.75).

Teachers and principals do not agree that students' scores on on-line high stakes tests provide feedback for schools to improve the curriculum (teacher mean = 3.66, principal mean = 1.0) and that the on-line high stakes test content is aligned with the school's curriculum (teacher mean = 1.83, principal mean = 4.13).

Domain 2: Teaching

The second set of five questions, 6-10, in the survey mapped to the teaching domain in both the teacher and principal surveys. The teaching domain captured teacher's thoughts about the relationship between the teacher's pedagogy and the student's performance on the high stakes assessment (Table 3).

Table 3

Mean and Standard Deviation for Questions in the Teaching Domain by Teacher

Item #	**Question**	**School A (N = 45)**		**School B (N = 15)**		**Combined (N = 60)**	
		M	SD	M	SD	M	SD
6	Students' scores on on-line high stakes test are a valid measure of teaching ability.	2.64	2	3.46	0.74	3.53	0.74
7	The quality of teachers' instruction is directly related to student performance on a on-line high stakes test.	4.15	4.5	4.13	0.35	4.15	0.36
8	On-line high stakes testing requires test preparation that diminishes time to teach subject content.	2.64	3.75	2.3	0.51	2.56	0.62
9	On-line high stakes testing has increased cooperation among teachers.	2.17	1	3.06	1.27	2.4	1.12
10	On-line high stakes testing has increased teacher and principal cooperation.	3.11	2.75	3.8	0.56	3.28	0.76

School A analyses found teachers agree to being undecided on the student's scores being a valid measure of their teaching ability (M = 2.64), on-line high stakes testing

requires test preparation that diminishes time to teach subject content (M = 2.64), and on-line high stakes testing has increased cooperation among teachers (M = 2.17). Teachers do not agree that instruction is directly related to student performance on a high stakes test. (M = 4.15) or on-line high stakes testing increasing teacher and principal cooperation. (M = 3.11). School B teachers are in agreement with School A faculty on 3 of the 5 survey questions (7, 8, and 10).

Table 4

Mean and Standard Deviation for Questions in the Teaching Domain by Principals

Item #	**Question**	**School A (N = 4)**		**School B (N = 4)**		**Combined (N = 8)**	
		M	SD	M	SD	M	SD
6	Students' scores on on-line high stakes test are a valid measure of teaching ability.	2.0	0.0	4.25	0.5	4.37	0.52
7	The quality of teachers' instruction is directly related to student performance on an on-line high stakes test.	4.5	0.57	1.7	0.5	1.87	0.35
8	On-line high stakes testing requires test preparation that diminishes time to teach subject content.	3.75	0.5	4.25	0.5	3.75	0.46
9	On-line high stakes testing has increased cooperation among teachers.	1.0	0.0	3.75	0.5	3.75	0.46
10	On-line high stakes testing has increased teacher and principal cooperation.	2.75	1.25	1.75	0.5	1.38	0.52

School A analyses found principals agree to strongly agree students' scores on on-line high stakes test are a valid measure of teaching ability. (M = 2.0), on-line high stakes testing has increased cooperation among teachers. (M = 1.0), and on-line high stakes testing has increased teacher and principal cooperation. (M = 2.75). Principals do not agree that the quality of teachers' instruction is directly related to student performance on a high stakes test. (M = 4.5) or on-line high stakes testing requires test preparation that

diminishes time to teach subject content (M = 3.75). School B principals are in agreement with School A faculty on 1 of the 5 survey questions (10).

Overall, teachers and principals' perceptions range from agree to undecided on on-line high stakes testing requiring test preparation that diminishes time to teach subject content (teacher mean = 2.56, principal mean = 3.75) and that on-line high stakes testing has increased teacher and principal cooperation (teacher mean = 3.28, principal mean = 1.38).

Teachers and principals do not agree that the quality of teachers' instruction is directly related to student performance on a high stakes test (teacher mean = 3.28, principal mean = 1.38).

Domain 3: Stress

The third set of five questions, 11-15, in the survey mapped to the stress domain in both the teacher and principal surveys. The stress domain captured teacher's thoughts about the relationship between the Georgia Milestones Assessment and the stress that teacher's feel during testing (Table 5).

Table 5

Mean and Standard Deviation for Questions in the Stress Domain by Teachers

Item #	Question	School A (N = 45)		School B (N = 15)		Combined (N = 60)	
		M	SD	M	SD	M	SD
11	On-line high stakes testing increases when the school receives a failing grade.	3.04	0.63	3.06	0.25	3.05	0.56
12	Teachers experience stress in the effort to maintain their school's accountability grade.	1.17	0.38	1.2	0.41	1.18	0.39
13	District supervisors' pressure to improve high stakes test scores increases teacher stress.	1.02	0.14	1.06	0.25	1.03	0.18
14	Principals' pressure to improve high stakes test scores increases teacher stress.	1.35	0.48	1.13	0.35	1.3	0.46
15	Punitive measures associated with high stakes testing increase teachers stress.	1.55	0.75	1.86	0.51	1.63	0.71

School A analyses found teachers agree to strongly agree that they experience stress in the effort to maintain their school's accountability grade (M = 1.17), district supervisors' pressure to improve high stakes test scores increases teacher stress (M = 1.02), principals' pressure to improve high stakes test scores increases teacher stress (M = 1.35), and punitive measures associated with high stakes testing increase teachers stress (M = 1.55). Teachers do not agree that on-line high stakes testing increases when the school receives a failing grade (M = 3.04). School B teachers are in agreement with School A faculty on 5 of the 5 survey questions (11, 12, 13, 14, and 15).

Table 6

Mean and Standard Deviation for Questions in the Stress Domain by Principals

Item #	Question	School A (N = 4)		School B (N = 4)		Combined (N = 8)	
		M	SD	M	SD	M	SD
11	On-line high stakes testing increases when the school receives a failing grade.	3.0	0.8	3.0	0.81	3.0	0.76
12	Teachers experience stress in the effort to maintain their school's accountability grade.	1.75	0.5	2.0	0.0	1.87	0.35
13	District supervisors' pressure to improve on-line high stakes test scores increases teacher stress.	2.0	0.0	2.0	0.0	2.0	0.0
14	Principals' pressure to improve on-line high stakes test scores increases teacher stress.	2.5	0.5	2.0	0.0	1.87	0.35
15	Punitive measures associated with on-line high stakes testing increase teachers stress.	4.25	0.5	4.0	0.81	4.13	0.64

School A analyses found principals agree to strongly agree teachers' experience stress in the effort to maintain their school's accountability grade (M = 1.75), district supervisors' pressure to improve high stakes test scores increases teacher stress (M = 2.0), and principals' pressure to improve high stakes test scores increases teacher stress (M = 2.5). Principals do not agree on-line high stakes testing increases when the school receives a failing grade (M = 3.0) or punitive measures associated with high stakes testing increase teachers stress (M = 4.25). School B administrators are in agreement with School A faculty on 5 of the 5 survey questions (11, 12, 13, 14, and 15).

Overall, teachers and principals agree that teachers' experience stress in the effort to maintain their school's accountability grade (teacher mean = 1.18, principal mean = 1.87), district supervisors' pressure to improve high stakes test scores increases teacher stress (teacher mean = 1.03, principal mean = 2.0), and principals' pressure to improve

high stakes test scores increases teacher stress (teacher mean = 1.3, principal mean = 1.87).

Teachers and principals do not agree that punitive measures associated with high stakes testing increase teachers stress (teacher mean = 1.63, principal mean = 4.13)

Domain 4: Accountability

The last three questions, 16-18, in the survey mapped to the curriculum domain in both the teacher and principal surveys. The curriculum domain captured teacher's thoughts about the relationship between the Georgia Milestones Assessment and teacher accountability (Table 7).

Table 7

Mean and Standard Deviation for Questions in the Accountability Domain by Teachers

Item #	Question	School A (N = 45)		School B (N = 15)		Combined (N = 60)	
		M	SD	M	SD	M	SD
16	On-line high stakes testing has increased teachers' awareness of the accountability issues in education.	2.4	0.53	1.13	0.35	2.08	0.74
17	On-line high stakes testing is a reform measure that improves the quality of education.	4.68	0.46	4.86	0.35	4.73	0.44
18	Teachers are more accountable because of on-line high stakes testing.	4.0	0.63	4.13	0.35	4.03	0.58

School A analyses found teachers agree to strongly agree high stakes testing has increased teachers' awareness of the accountability issues in education (M = 2.4). Teachers do not agree high stakes testing is a reform measure that improves the quality of education (M = 4.68) or teachers are more accountable because of on-line high stakes testing (M = 4.0). School B teachers are in agreement with School A faculty on 3 of the 3 survey questions (16, 17, and 18).

Table 8

Mean and Standard Deviation for Questions in the Accountability Domain by Principals

Item #	Question	School A (N = 4)		School B (N = 4)		Combined (N = 8)	
		M	SD	M	SD	M	SD
16	On-line high stakes testing has increased teachers' awareness of the accountability issues in education.	1.0	0.0	1.0	0.0	1.0	0.0
17	On-line high stakes testing is a reform measure that improves the quality of education.	4.0	0.0	4.0	0.0	4.0	0.0
18	Teachers are more accountable because of on-line high stakes testing.	4.0	0.0	4.0	0.0	4.0	0.0

School A analyses found principals agree to strongly agree high stakes testing has increased teachers' awareness of the accountability issues in education (M = 1.0). Principals do not agree high stakes testing is a reform measure that improves the quality of education (M = 4.0) or teachers are more accountable because of on-line high stakes testing (M = 4.0). School B principals are in perfect agreement with School A faculty on 3 of the 3 survey questions (16, 17, and 18).

Summary

Overall, teachers and principals agree that high stakes testing has increased teachers' awareness of the accountability issues in education (teacher mean = 2.08, principal mean = 1.0). Teachers and principals are in agreement that they disagree that on-line high stakes testing is a reform measure that improves the quality of education (teacher mean = 4.73, principal mean = 4.0) and teachers are more accountable because of on-line high stakes testing (teacher mean = 4.03, principal mean = 4.0).

Teachers and principals in both schools are in agreement on all of the survey questions (16, 17, and 18) within the Accountability domain.

Open-Ended Question Findings

The same four questions were asked on the teacher and principal surveys. Each question was analyzed for common domain/themes by teacher and principal (Appendix E). Findings are presented by question.

Question 1: What is your opinion of on-line testing?

Three (3) teacher domains/themes were found when analyzing the data. Themes were: (1) Dislike of on-line testing (38%),(2) Quick return of scores that could be used for remediation (15%), and (3) Meeting the learning styles of students (25%). A typical teacher response for theme 1 was: *I dislike* o*n-line testing because it does not reflect what the students have learned throughout the school year.* A typical teacher response for theme 2 was: *Scores returned sooner which allows more time for planning, remediation, retentions.* A typical teacher response for theme 3 was: *Meets learning styles needs of students (visual, technology).*

Three (3) principal themes were found when analyzing the data. Themes were: (1) Less paper work (37%), (2) Quick return of scores (37.5%), and (3) Less possibility for cheating (25%). A typical principal response for theme 1 was: *Less paperwork.* A typical principal response for theme 2 was: *Scores are returned quickly, and data can be used for remediation.* A typical teacher response for theme 3 was: *Online reduces the possibility for cheating to occur during testing.* Teachers and principals agreed on 1 out the 3 themes (Quick return of scores).

The teachers and principals' responses varied greatly for this question. Most responses centered on the ease and quickness of receiving scores and providing remediation for students to the state's response to meeting the varied learning styles that

exists within the classroom (differentiated learning). There were a large number (25) of teachers that expressed their dislike for on-line testing.

Open-ended Question 2: What is your opinion of the accuracy of on-line high stakes testing?

Three (3) teacher themes were found when analyzing the data. Themes were: (1) Scores do not accurately reflect what students have learned throughout the year (46.7%), (2) Scores should not be used to evaluate teachers (23.3%), and (3) Computer glitches and typing skills may have an effect on student scores (30%). A typical teacher response for theme 1 was: *The test scores do not give an accurate accounting of all the students have learned throughout the year.* A typical teacher response for theme 2 was: *Teachers evaluations should not be based on the scores of the students*. A typical teacher response for theme 3 was: *Computer glitches and typing skills play a large part in accurately accessing the students' mastery of content when taking on-line tests.*

Three (3) principal themes were found when analyzing the data. Themes were: (1) Reduces cheating (50%), (2) Reflects how well the teachers are teaching the curriculum (37.5%), and (3) Helps to determine areas of need (12.5%). A typical principal response for theme 1 was: *On-line testing helps to reduce cheating among students.* A typical principal response for theme 2 was: *On-line testing can help determine how well the teachers taught the state curriculum and can be used to better evaluate teachers*. A typical principal response for theme 3 was: *Test scores can be used to better determine the needs of students, as well as, areas of strength and weakness of teachers.* Teachers and principals did not agree on any of the themes.

The teachers and principal's thoughts regarding question two were contradictory.

The teachers felt that the scores on on-line tests are not an accurate indicator of a student's mastery of actual information outlined within the curriculum and cannot be tested while principals felt that on-line high stakes testing reflected how well the teachers are teaching the curriculum.

Open-ended Question 3: What are barriers you see to on-line high stakes testing?

Three (3) teacher themes were found when analyzing the data. Themes were: (1) Too many computer glitches (30%), (2) Limited student training on using the computers (41.7%), and (3) Student scores may be affected by the computer glitches (28.3%). A typical teacher response for theme 1 was: *Technical glitches (computer signing off automatically and having to login again).* A typical teacher response for theme 2 was: *Limited keyboarding knowledge for significant number of students.* A typical teacher response for theme 3 was: *Technology glitches interfere with time allotted to take test.*

Three (3) principal themes were found when analyzing the data. Themes were: (1) Technology problems (12.5%), (2) Student's familiarity with keyboard usage (12.5%), and (3) Scheduling of students to take the test with limited number of computers (75%). A typical principal response for theme 1 was: *Technology problems occurred throughout the entire testing process.* A typical principal response for theme 2 was: *Very few students knew how to type*. A typical principal response for theme 3 was: *Limited technology devices (scheduling, problems for accommodating population).* Teachers and principals agreed on 2 out the 3 themes (Technology issues and student's limited keyboarding skills).

One of the main barriers that all schools noted, teachers and principals were the glitches in technology experienced by the students. Although the glitches may have been

corrected within a quick response time, those glitches had an effect on the students and possible the scores.

Open-ended Question 4: How do you feel on-line high stakes testing has changed instruction?

Three (3) teacher themes were found when analyzing the data. Themes were: (1) requires more usage of computer programs during instruction (58.3%), (2) requires students to learn keyboarding skills (36.7%), and (3) requires more professional development for teachers on incorporating computers in instructional practices (5%). A typical teacher response for theme 1 was: *Teachers are allowing students to participate in more on-line programs.* A typical teacher response for theme 2 was: *Schools are investing in more web-based programs to learn typing skills.* A typical teacher response for theme 3 was: *Principals will have to provide more professional learning on computer usage in the classroom.*

Three (3) principal themes were found when analyzing the data. Themes were: (1) increase usage of computer-based programs in the classroom (12.5%), (2) students' need for learning keyboard skills (37.5%), and (3) more training for teachers on using computers in the classroom (58%). A typical principal response for theme 1 was: *The student's need to interact with computers more in the classroom.* A typical principal response for theme 2 was: *Students must learn to type to be successful during on-line testing*. A typical principal response for theme 3 was: *Teachers are going to need more professional development of effectively using computers in the classroom.* Teachers and principals agreed on all of the themes.

One of the main themes that teachers and principals noted from all schools was

the need for students to learn typing skills and have more interaction with using computers in the classroom. Lack of typing skills had an effect on the amount of time allotted for students to answer questions and may have affected their scores, as well.

Overall teachers and principals thought that the on-line testing of the Georgia Milestones had its benefits, however, they noted a few things that state officials may not have considered when creating the state testing implementation plan. Teacher and principals noted that testing on-line was more secure, provided opportunities for remediation with its quick turn-around of providing student performance results back to the schools, and encouraged the increased usage of technology within the classroom environment. They also noted that computer glitches and students' typing skills may have an adverse effect on scores. Moreover, the majority of teachers and principals were in favor of using technology to administer on-line testing.

Summary

Tables 9 and 10 present domain means and standard deviations by teacher and principal by school and overall mean and standard deviation.

Table 9

Mean Scores and Standard Deviations by Domains for Teachers

Domain	School A (N = 45)		School B (N = 15)		Combined (N = 60)	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Curriculum	2.20	1.12	2.03	1.03	2.15	1.10
Teaching	2.93	1.10	3.36	0.98	3.19	0.99
Stress	1.63	0.90	1.67	0.84	1.64	0.88
Accountability	3.70	1.11	3.37	1.67	3.61	1.27

Table 10

Mean Scores and Standard Deviations by Domains for Principals

Domain	School A (N = 4)		School B (N = 4)		Combined (N = 8)	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Curriculum	1.96	1.28	1.9	1.65	1.9	1.17
Teaching	3.15	1.5	3.15	1.27	3.15	1.37
Stress	2.55	1.10	2.6	0.94	2.57	1.01
Accountability	3.0	1.48	3.0	1.48	3.0	1.45

Overall analyses by domain, where the school's findings and teachers and principals' findings were combined, can be found in Tables 9 and 10. According to the analyses, teachers (M = 2.15) and principals (M = 1.9) agree to strongly agree that within the Curriculum domain (M = 2.01), the Georgia Milestones Assessment reflects the state curriculum. Both groups, teachers (M = 1.64) and principals (M = 2.57), also agree to strongly agree that the Georgia Milestones Assessment is a contributing factor to the level of stress that teachers' feel within the Stress domain (M = 1.90). Both groups do not agree on the role that on-line testing plays within the Teaching (M = 3.18) and Accountability (M = 3.54) domains. Overall, all of the participants agreed on 2 out of the 4 domains (Curriculum and Stress) surveyed.

Chapter Summary

Chapter IV presented an analysis of data gathered from the responses of 3-5 elementary school teachers and principals regarding their perceptions of Georgia Milestones Assessment testing (i.e. on-line high stakes testing). The major findings of the study regarding teacher perceptions of on-line high stakes testing were that School A and School B teachers' perceptions differ on four of eighteen survey questions, however

overall perceptions of on-line high stakes testing were about the same. The teachers and principal's perceptions slightly differed by domain with the stress domain showing the greatest difference in perception among both groups. Responses of the open-ended questions provide the qualitative analysis of the findings from each group. Whereas 44% of the participants noted that on-line testing is more secure than paper pencil testing, 16% of the participants thought that technology glitches had an effect on the school's scores. Additionally, 17.6% of the participants thought that receiving scores back quickly and dealing with less paperwork (23.5%) were contributing factors to supporting the continued use of on-line testing.

CHAPTER V

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

This study focused on determining grades 3 – 5 teachers' and principals' perceptions regarding Georgia Milestones on-line testing in a Georgia school district. As a response to the No Child Left Behind Act of 2001 and later, Every Child Succeeds Act signed in 2015, states are continuing the task of improving the testing environment. The current study was designed to accomplish two goals. The first was to examine how grades 3-5 teachers and principals perceive Georgia Milestones on-line testing in a Georgia school district and second, to examine the perceptions through the lens of the domains of Curriculum, Teaching, Stress and Accountability.

Summary of Study

The curriculum domain captured teacher's thoughts about the relationship between the Georgia Milestones Assessment and the state curriculum. The results from this study indicated that the Georgia Milestones on-line testing format has made teachers reassess their beliefs about what is important to teach; however, the results indicated that teachers thought that the test was not aligned with the state curriculum nor could the results be used to improve the school's instructional practices. The principal's perspective on on-line testing was slightly different regarding the curriculum being used to improve instructional practices and its alignment with the curriculum. They did not agree that on-line testing changed teacher's beliefs about curricula content that is important to teach.

The teaching domain captured teacher's thoughts about the relationship between the teacher's pedagogy and the student's performance on the high stakes assessment.

Teachers and principals were not in agreement about the effect that on-line testing had on the quality of instruction. Teachers did not think the Georgia Milestones test reflected on the quality of instruction and that student scores should not be used as an evaluative measure. Principals, on the other hand, did not agree with the teachers' perspective. They thought that the students' scores on the Georgia Milestones assessment reflected the teacher's quality of teaching. However, within the open-ended section of the survey, both groups noted that technology glitches and students' typing skills may have an effect on the students' scores.

The stress domain captured teachers' thoughts about the relationship between the Georgia Milestones Assessment and the stress that teachers feel during testing. Stress is one of the main causes of the teacher exodus that is happening all over the country. Teachers feel over-worked and underpaid and with the added stress of high stakes testing, more teachers are not seeing the benefits of staying in the education career field for thirty plus years. The results in this study indicated that teachers and principals agree that teachers' experience stress in the effort to maintain their school's accountability grade, trying to improve high stakes test scores increases and from principals' pressure to improve high stakes test scores increases teacher stress. Teachers' notes in the open-ended response section that technology glitches during the testing period also adds stress to the entire testing environment for students and staff members.

The accountability domain captured teachers' thoughts about the relationship between the Georgia Milestones Assessment and teacher accountability. The study indicated that on-line high stakes testing has made teachers more aware of accountability

issues within the field of education. There was no agreement on the actual effect that accountability plays with high stakes testing.

Discussions

Discussions on educational testing remains focused upon how the teachers and principals seem to return to a common theme: The love of learning and how to promote this in a transition period for educational reform. With so many governmental policies in place to guide the standards for what should be taking place in the classroom, there is little wonder why teachers are concerned, anxious, and stressed about the performance of his or her students on the Georgia Milestones on-line testing. Principals seem to have a broader view of the situation but still feel the need to address how the learning environment is changing to include advanced on-line e-learning tools, in the hopes of enhancing and bringing value to the classroom for all learners. One cannot ignore the impact that the teaching experience and the experience principals have in their new roles places on the stress levels of teachers. However, the roles create gaps instead of forging collaboration. The issue remains, according to the findings in this study, in the stress level of the teacher as the principal expects the teacher to take upon the burden of testing outcomes being directly related to his or her instruction and quality of implementing the curriculum. The little support they feel from the school may be indicative of the teaching personality of “my classroom, my way;” also, the traditional educational format that does not fit with the Georgia Milestones on-line testing process. Yet, the state wants the school to adopt the Georgia Milestones online testing format causing another gap to form.

This study supported the literature that seeking an on-line standard for testing for students that are already affected by the many inconsistencies defined by gaps and disparities of socio-economic levels is a credible discussion topic for educators. This issue is found in communities identifying as rural or urban with high crime rates and lack of parental support within the education arena. It seems that this disparity knows little discrimination but in fact seems to widen the gap as more social interaction is reliant upon technological tools. In order to narrow gaps, address disparities, states like Georgia must better align their vision of the future for education with actual actions that seek better ways to provide funding and investments in community infrastructures to address the need for increased support of teachers. To add value at this point would mean not only allowing all schools the same advantage, but principals would need to offer every teacher the same level of support system.

The Georgia Milestones on-line testing system our state seeks to implement as a standard format for testing is ambitious and may benefit the educational system but by the same token, it presents a number of challenges in terms of how to align standards (curriculum) with needs to address (teaching) and alleviate the stress teachers feel at the classroom level. One could argue such needs begin at home with the community directly. Does the problem of gaps and disparities belong to the community it directly impacts? Or really does the problem belong to all of us, even those of us living beyond Georgia's borders? This issue of education reform and Georgia Milestones on-line testing, also, points to the greater issue of how education falls short of representing great value as a tool for all students in America and not just a few.

The issue is that not every person has the same access or need of access. There are

the concerns that with too much focus upon how e-learning can be used to solve some of the gaps existing today, it also creates another disparity for what pure knowledge is and the value it may have at the core of the system. Several researchers noted that if evaluations are done in the on-line format, what sort of accuracy and true depiction that serves to the student and the school in terms of further designing the learning environments. This places accountability for the score in the hands of the state as it also sets the standard for how the on-line testing environment and protocols will be implemented. The thought here is standards will not be uniform because the standards set by Common Core offers so much flexibility. This in turn creates anxiety for the teacher and new burdens for the principal in a way that cannot be defined in traditional educational constructs. While the unknown is exciting in many ways to face the challenge for untapped opportunities for learning in this transitional period, there is also the concern for long term impact upon the mental health of educational professionals.

Implication for transitioning over to the use of more e-tools in educational environments points to not only the anxiety, stress, and concern for accountability but also a division of those who believe technology will be the proven solution and those who do not believe it solves enough problems to warrant such wide spread adoption of Georgia Milestones on-line testing. Place these two camps in the same school and there will be conflict. Yet the implication exists that the standards for human social interaction has changed to a point where people rely more completely upon the handheld device and children born of this era will not know any other context of communication other than digital smart devices. Furthermore, researchers argue the limitless capacity of how the Georgia Milestones on-line tool and the handheld device increasing one’s ability to

thinking critically and creatively. This may allow the student to gain a better balance of subject matter that they are good at and achievement toward well rounded skills and hobbies which others commend as desirable and advantageous to the modern complex world. This can serve the student as an individual needed as a leader in business, innovation, and policy, just to name a few. To use the learning device as a means to learn music and art, seek research examples for biology a project; this can only extend the knowledge base when there is access.

It is noted that there is the implication that 'too much of a good thing' is actually bad. The notion that the student can find entertainment value as well as educational application on the same device with similar on-line content also may not be relevant to the student's context or ease of usefulness. There is the concern that not all students have the same level of exposure to on-line testing and therefore use of on-line testing as a scoring mechanism may not benefit those students who lack computer skills such as with the keyboard or protocols for logging in or out. These differences may be found in correlation with socioeconomic conditions but also the teachers find these mismatches of skills as distracting and creating conflict that is unneeded.

Finally, one cannot ignore how powerful the opportunity created by the on-line learning device to integrate access to knowledge. However, limits found within some rural and urban communities are defined by the level of access to these tools and the amount of knowledge users have about them. On-line Georgia Milestones testing can create a path toward narrowing the gap, but this is only achievable when conflict is not a byproduct of the environment. The learning device, the e-tool or e-learning environment

has great potential to integrate formats that increases critical thinking and collaboration, sharing of knowledge about the learning that has not been applicable before to the traditional classroom. Possibly the amount of unknown opportunities also challenges teachers with access but for those who lack access and have little infrastructure, these classrooms and the children there are still at a marked disadvantage.

Conclusions

Findings in this study are similar to prior research regarding teacher's perceptions of high stakes testing; however, analyses found differences existed between this study and research. Similarities, including the increase of the use of technology in the classroom, were limited to teacher's perceptions in improvements in student achievement, teaching and learning in the classroom, and accurately reflecting expectations of the common core standards. This study's findings seem to support the views expressed in a study conducted by Rentner, Kober, Ferguson, McMurrer, and Frizzell (2014) that technology issues created a major challenge when administering the on-line assessments. Thereby adding a higher level of stress in the overall school environment and highlighted the disparity of the student's use of technology outside of the school environment (Kornhaber, Griffith, & Tyler, 2014). Nevertheless, test security and quick access to scores were considered as a positive reason for continuing the use of on-line testing.

Additionally, survey analyses found that teachers and principals, in the curriculum domain, believed that on-line high-stakes testing led teachers to reassess beliefs about subject matter that is important to teach. However, only principals believed that the test scores provided feedback to improve curriculum. The survey also revealed, in the

teaching domain, that teachers and principals have different perceptions on the quality of teachers' instruction being directly related to student performance on on-line high-stakes tests. The survey analyses of the stress and accountability domains revealed that improving on-line high stakes scores and accountability grades aided in the stress that teachers experience, however, the presence of on-line testing has increased teachers' awareness of accountability measures. Open-ended question analyses found both teachers and principals believe that technology glitches and the student's limited keyboarding skills had an effect on the students' overall test scores. Moreover, on-line high stakes testing has helped in increasing technology usage in the classroom; however, the limited student access to this technology may have an effect on their performance on these tests.

Implications

The data focuses primarily on the four domains of curriculum, teaching, stress and accountability. All of these domains impact teachers' perceptions of on-line testing, in one way or another, and ultimately shines a light on what happens in the classroom daily. Both groups, teacher and principals, shared pros and cons to the use of Georgia Milestones online tools but did not speculate about how its prolonged use may change the way students learn and apply knowledge to the future roads their lives take. Teachers stress the desirability to impact the student's life for the better and are actually truly hard upon themselves in terms of criticism about outcomes. Data suggests they stress about this not only because of fearing how a poor test score may reflect upon them under Common Core and teacher evaluations, but also how it may create a lack of value for education in our students. The teacher seeks the effectiveness of one on one interaction of

the class and fears how the Georgia Milestones on-line learning environment changes this level of intimacy. The Georgia Milestones on-line format creates isolation of the student that many students already feel especially in socioeconomic despaired sections of the community. The wider implication here is how such facets of social norms like value for technological tools and advancements serve the community at large but also for some change the expectation of having knowledge sources readily available in the form of entertainment.

Teachers new to the professional can use the findings from this study to get an idea of the issues that experienced teachers and students face during on-line high-stakes testing. The teachers will then be able to align their teaching pedagogy and activities to provide more student interactions with technological devices and web-based activities to help students to be at ease with its use during testing. This information will also assist the principals and the district personnel (i.e. superintendent) with understanding the on-line testing environment and what difficulties teachers and students face during this period of time. Hopefully, the principals and superintendent will seek to address the issues of more student interactions with technological devices, as well as, how to schedule testing in a way that results in the least amount of technological glitches during the testing period.

The Georgia Department of Education (GADOE) and educational researchers can benefit from the analyses found in this study to articulate what difficulties occur throughout the state during the period of on-line testing. The implications from this study highly suggest that technological glitches and limited typing skills affect student performances, thereby causing a ripple effect on teacher stress, teacher turnover rates, student dropout rates and other issues that affect teacher and students' beliefs about on-

line high-stakes testing. The research found in this study will also help the GADOE and researchers understand how teachers feel about the correlation between the on-line test and the curriculum being taught. Further investigations, based on the findings from this study, may help with the development of an assessment that truly reflects curriculum, teaching, and accountability while alleviating stress from the overall school environment.

Recommendations

Predicated on the findings and conclusions of this study, more in-depth research is necessary regarding the perceptions of Georgia on-line testing on all teachers, students, and achievement. The following implications for research are offered for future studies:

1. This study should be replicated with an emphasis of analyzing demographics of new teachers versus experienced teachers and/or male teachers versus female teachers who teach subjects that will be tested. This study will provide perceptions from the male point of view and the female point of view while analyzing the thoughts and perceptions of new teachers versus experienced teachers. This study will help principals to get an understanding of the nature of the responses and the perspective from which they arise. New teachers and males' outlook on testing may reflect differently than those of experienced teachers and few males which mainly dominate a school. This study will allow the various voices within a school to be highlighted.

2. This study should be expanded to include students' perceptions. This study will help to determine the impact that on-line testing has on students.

3. A new study should be created to determine whether on-line testing impacts middle and high school students. This study will help administrators to get an understanding of the long-term effect that on-line high stakes testing has on students as

they matriculate through school.

Impact Statement

This study has provided me with a lot of insight into the thoughts that teachers and principals have regarding the Georgia Milestones Assessment. I was very surprised to discover that there are a large percentage of teachers that do not like testing at all. As a teacher, we know the importance of testing and the benefits that the data provides in helping to ensure that our students master the curriculum in which we teach. I thought that, like me, every teacher value the information that can be gained from the test results when it comes to improving the education of our students. I was also surprised to discover the concern about test scores that teachers expressed due to the glitches in the technology. I, like the majority of the world, thought that the use of technology in the testing environment would help relax the students and provide a creative way to administer the test. Unfortunately, the technology glitches cause the teachers and students anxiety levels to rise which makes the administering of the test more stressful then anyone could have imagined.

Unfortunately, the data from the study did not surprise me when it highlighted the differences in perspective of teachers and principals within in the teaching domain. Principals have always wanted to use more stringent means for evaluating teachers, and the Georgia Milestones Assessment has provided the information that they need in order to "measure" a teacher's ability to teach their students. Within the teaching domain, teachers and principals did not agree that the quality of teachers' instruction is directly related to student performance on a high stakes test. Many principals believe that

you can determine how well a teacher is teaching based on the students' scores. These results again highlighted how administrators forget what it was like to teach once they get out of the classroom.

This entire study has caused me to re-evaluate how I think about testing and how I work with my grade-level team as we prepare our students for the Georgia Milestones Assessment. I have learned from this study that it is extremely important for the teachers on my grade-level and the other grade-level teams to band together to assist each other through the testing process so that we do not feel overwhelmed and stressed out as we prepare our students for the testing period. Additionally, this study has made me more aware of the need and importance of providing my students with the opportunity to prepare for the on-line test by providing more opportunities for my students to learn keyboarding skills and deal with glitches that may arise when using technology. This practice will help the students to remain calm when situations occur during the high-stakes on-line testing period, as well as, prepare me to better assist my students when and if it happens during the Georgia Milestones Assessment. Also, the results from this study has highlighted the need for the faculty and administration to work together more closely on a consistent basis to ensure that we are all aware of issues that may arise during testing and to develop a plan of action to better deal with these issues so that we all are not stressed during this time.

REFERENCES

Arroyo, I., Beal, C., Murray, T., Walles, R., & Woolf, B. (2004, January). Web-based intelligent multimedia tutoring for high stakes achievement tests. *Intelligent Tutoring Systems*, pp. 468-477.

Assessment System [Def. 1]. (n.d.). *Education Dictionary Online.* In mondofacto. Retrieved June 11. 2017, from www.mondofacto. Retrieved June 11, 2017, from www.momdofacto.com/facts/dictionary?query=assessment=system&action=look+it+up

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bathon, J. (2013). For districts, online testing has legal liabilities. *THE Journal*, *40*(7), 17-20.

Boevé, A. J., Meijer, R. R., Albers, C. J., Beetsma, Y., & Bosker, R. J. (2015). Introducing computer-based testing in high-stakes exams in higher education: Results of a field experiment. *PLoS ONE, 10*(12), e0143616. doi:10.1371/journal.pone.0143616

Boser, U., Baffour, P., & Vela, S. (2016). *A look at the education crisis: Tests, standards, and the future of American education.* Washington, DC: Center for American Progress. Retrieved from http://www.americanprogress.org

Boser, U., & Brown, C. (2016). *Lessons from state performance on NAEP: Why some high-poverty students score better than others.* Washington, DC: Center for American Progress. Retrieved from http://www.americanprogress.org

Brockmeier, Lantry L.; Pate, James L.; and Leech, Don W. (2008) "Validation of the Principal's . High Stakes Testing Survey," *Georgia Educational Researcher*: Vol. 7 : Iss. 2 , Article 4.

Burks, B. A., Beziat, T. R., Danley, S., Davis, K., Lowery, H., & Lucas, J. (2015). Adapting to change: Teacher perceptions of implementing the Common Core Standards. *Education, 136,* 253-258

Burnette, D. (2016). Testing stumbles prompt legislation in affected states. *Education Weekly, 35*(23), 16.

Camera, L. (2016a). Online exams' rough rollout piques anxiety. *Education Weekly, 35*(23), 17.

Camera, L. (2016b). Tennessee district perseveres amid online-testing woes. *Education Weekly, 35*(23), 14-15, 17.

Cassady, J. C. & Gridley, B. E. (2007). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning, and Assessment, 4*(1). Retrieved from http://www.jtla.org

Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior, 29,* 1889-1895. doi:10.1016/j.chb.2013.03.008

Clarke, M. M., Madaus, G. F., Horn, C. L., & Ramos, M. A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies, 32,* 159-181.

Colwell, N. M. (2013). Test anxiety, computer-adaptive testing and the Common Core. *Journal of Education and Training Studies, 1*(2), 50-60. doi:10.11114/jets.v1i2.101

Conley, D. T. (2015). A new era for educational assessment. *Education Policy Analysis Archives, 23*(8). http://dx.doi.org/10.14507/epaa.v23.1983

Cooper, D. (2011). Talk *about assessment: High school strategies and tools.* Toronto, Canada: Nelson Education.

Darling-Hammond, L. (1990). Instructional policy into practice: "The power of the bottom over the top." *Educational Evaluation and Policy Analysis, 12,* 339-347.

Darling-Hammond, L. (2014, January/February). Testing to, and beyond, the Common Core. *Principal*, pp. 9-12.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15.

Darling-Hammond, L., & Falk, B. (2013). *Teacher learning through assessment.* Washington, DC: Center for American Progress. Retrieved from http://www.americanprogress.org

Davis, L. L., Strain-Seymour, E., & Gay, H. (2013). *Testing on tablets: Part II of a series of usability studies on the use of tablets for K-12 assessment programs* (White Paper). Pearson. Retrieved from http://www.pearsoned.com/

Davis, M. R. (2014). Online testing glitches causing distrust in technology. *Education Weekly, 33*(30), 20.

Debuse, J. W., & Lawley, M. (2016). Benefits and drawbacks of computer-based assessments and feedback systems: Student and educator perspectives. *British Journal of Educational Technology, 47,* 294–301. doi:10.1111/bjet.12232

Dee, T. S., & Jacob, B. A. (2010, Fall). The impact of No Child Left Behind on students, teachers, and schools. *Brookings Papers on Economic Activity,* No. 2, 149-207.

Delen, E. (2015). Enhancing a computer-based testing environment with optimum response time. *Eurasia Journal of Mathematics, Science & Technology Education, 11,* 1457-1472. doi:10.12973/eurasia.2015.1404a

DePascale, C., Dadey, N., & Lyons, S. (2016). *Score compatibility across computerized assessment delivery devices.* Council of Chief State School Officers. Retrieved from http://www.ccsso.org/Documents/CCSSO%20TILSA%20Score%20Comparability%20Across%20Devices.pdf

Desimone, L. M. (2013). Reform before NCLB. *Phi Delta Kappan, 94*(8), 59-61.

Doorey, N., & Polikoff, M. (2016). *Evaluating the content and quality of next generation assessments.* Thomas B. Fordham Institute. Retrieved from https://edex.s3-us-west-2.amazonaws.com/(02.09%20-%20Final%20Published)%20Evaluating%20the%20Content%20and%20Quality%20of%20Next%20Generation%20Assessments.pdf

Doran, L. (2016). Tennessee online test crashes, causing return to paper and pencil. *Education Weekly, 35*(21), 7.

Duffy, M., Giordano, V. A., Farrell, J. B., Paneque, O. M., & Crump, G. B. (2008). No Child Left Behind: Values and research issues in high-stakes assessments. *Counseling and Values, 53*, 53-66.

EdTech Strategies. (2015). *Pencils down: The shift to online & computer-based testing.* Retrieved from https://www.edtechstrategies.com/research-and-writing/usk-8-testing/

Education Trust (2015). http://www.edtrust.org/dc/about

Explorable.com https://explorable.com/convenience-sampling

Fennell, M. (2016). What educators need to know about ESSA. *Educational Leadership, 73*(9), 62-65.

Fink, J. W. (2015, Winter). Assessing the new assessments. *Scholastic Instructor,* pp. 26-28.

Fletcher (2011) Fletcher, D. (2009, December 11). A Brief History of Standardized Testing.fromhttp://content.time.com/time/nation/article/0,8599,1947019 ,00.html\

Gewirtz, C. (2013). Transition to online tests sparks fears. *Education Weekly, 33*(10), 1, 15.

Gewirtz, C. (2015). Online testing drives “tech prep” priorities. *Education Weekly, 34*(30), 1, 12.

Georgia Department of Education. (2014). *Teacher Keys Effectiveness System fact sheets.* Retrieved from https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/FY15%20TKES%20and%20LKES%20Documents/2014-2015%20TKES%20Fact%20Sheets%20-formatted%206-26-14.pdf

Georgia Department of Education. (2016, July 26). *Georgia Milestones scores increase for most subjects.* Retrieved from http://www.gadoe.org/External-Affairs-and-Policy/communications/Pages/PressReleaseDetails.aspx?PressView=default&pid=454.

Georgia Department of Education. (2017). *Georgia Milestones Assessment System.* Retrieved from http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/Georgia-Milestones-Assessment-System.aspx

Guisbond, L., Neill, M., & Schaeffer, B. (2012). *NCLB's lost decade for educational progress: What can we learn from this policy failure?* Jamaica Plain, MA: FairTest. Retrieved from http://fairtest.org/NCLB-lost-decade-report-home.

Hall, A. H., Hutchison, A., & White, K. M. (2015). Teachers' perceptions about the Common Core State Standards in writing. *Journal of Research in Education, 25*(1), 88-99.

Heiser, P., Simidian, G., Albert, D., Garruto, J., Catucci, D., Faustino, P.,…Caci, K. (2015, November). *Anxious for success: High anxiety in New York schools.* NYASP & New York State Boards Association. Retrieved from http://www.nyasp.biz/pdf_files/TestAnxietyReportFINAL.pdf

Herold, B. (2016a). Digital device choice could impact Common-Core test results, studies finding. *Education Weekly, 35*(36), 13.

Herold, B. (2016b). PARCC scores lower on computer exams. *Education Weekly, 35*(20), 1, 11.

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.1-23). Albany, NY: State University of New York Press.

Hopkins, W.G. (2008,July). *Quantitative Research Design.* Retrieved from http://sportsci.org/jour/0001/wghdesign.html

Hosseini, M., Abidin, M. J., & Baghdarnia, M. (2014). Comparability of test results of computer based tests (CBT) and paper and pencil tests (PPT) among English language learners in Iran. *Procedia - Social and Behavioral Sciences, 98*, 659 – 667. doi:10.1016/j.sbspro.2014.03.465

Huddleston, A. P., & Rockwell, E., C. (2015). Assessment for the masses: A historical critique of high-stakes testing in reading. *Texas Journal of Literacy Education, 3*(1), 38-49.

Jones, W. (2014). *Georgia Milestones to Challenge Students, Teachers*, 27 July 2017, onlineathens.com/local-news-k-12-schools/2014-07-26/georgia-milestones-challenge-students-teachers.

Kim, D.-H., & Huynh, H. (2007). Comparability of computer and paper-and-pencil versions of algebra and biology assessments. *Journal of Technology, Learning, and Assessment, 6*(4). Retrieved from http://www.jtla.org

Kornhaber, M. L., Griffith, K., & Tyler, A. (2014). It's not education by zip code anymore–but what is it? Conceptions of equity under the Common Core. *Education policy analysis archives*, *22*, 4.

Laughlin Davis, L., Strain-Seymour, E., & Gay, H. (2013). *Testing on tablets: Part II of a series of usability studies on the use of tablets for K-12 assessment programs* (White Paper). Pearson. Retrieved from https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/efficacy-and-research/schools/003__Testing-on-Tablets-Part-II_formatted.pdf

Lazarin, M. (2014). *Testing overload in America's schools.* Washington, DC: Center for American Progress. Retrieved from http://www.americanprogress.org

Lester, F. K. (2007). *Second handbook of research on mathematics.* Charlotte, NC: Information age.

Markow, D., Macia, L., & Lee, H. (2013). *The MetLife Survey of the American Teacher: Challenges for school leadership.* Retrieved from https://www.metlife.com/assets/cao/foundation/MetLife-Teacher-Survey-2012.pdf

Martin, M. (2016, August). School accountability systems and the Every Student Succeeds Act. *Re:Vision,* pp. 1-15. Durham, NC: Hunt Institute. Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED569952

McLaughlin, M., & Overturf, B. J. (2012). The Common Core: Insights into the K-12 Standards. *The Reading Teacher, 66,* 153–164. doi:10.1002/TRTR.01115

Murnane, R., & Papay, J. P. (2010). Teachers' views on No Child Left Behind: Support for the principles, concerns about the practice. *Journal of Economic Perspectives, 24*(3), 151-160. doi:10.1257/jep.24.3.151

Nelson, H. (2014). Testing more, teaching less: What America's obsession with student testing costs in money and lost instructional time. Retrieved from http://blogs.edweek.org/edweek/teadherbeat/TestingMore_TeachingLess.pdf

Ogletree, A., Ogletree, S., & Allen, B. (2014). Transition to online assessments: A personal perspective of meeting Common Core State Standards in an elementary school in Georgia, *Georgia Educational Researher,* 11(1), Article 7. doi: 10.20429/ger.2014.110107

Owens, S. J. (2015). *Georgia's teacher dropout crisis: A look at why nearly half of Georgia public school teachers are leaving the profession.* Atlanta: Georgia Department of Education. Retrieved from https://www.gadoe.org/External-Affairs-and-Policy/communications/Documents/Teacher%20Survey%20Results.pdf

Perception [Def.1]. (n.d.). *Merriam Webster Online.* In Merriam-Webster. Retrieved June 11, 2017. from http://www.merriam-websgter.com/dictionary/citation.

Poggio, J., Glasnapp, D. R., Yang, X., & Poggio, A. J. (2005). A comparative evaluation of score results from computerized and paper and pencil mathematics testing in a large scale state assessment program. *Journal of Technology, Learning, and Assessment, 3*(6). Retrieved from http://www.jtla.org

Polikoff, M. S. (2014). *Common Core State Standards assessments: Challenges and opportunities.* Washington, DC: Center for American Progress. Retrieved from http://www.americanprogress.org

Rennie Center. (2015, March). *Testing the test: A study of PARCC field trials in two school districts.* Retrieved from http://www.renniecenter.org

Rentner, D. D., Kober, N., Ferguson, M., McMurrer, J., & Frizzell, M. (2014). *Common Core State Standards in 2014: District implementation of consortia-developed assessments.* Washington, DC: Center on Education Policy. Retrieved from http://www.cep-dc.org

Riddle, M. (2015a, June1). *Principals and online testing—Part 1.* Retrieved from http://www.learningfirst.org/blog/principalsandonlinetestingparti

Riddle, M. (2015b, May 26). *Principals and online testing—Part 1I.* Retrieved from http://blog.nassp.org/2015/05/26/principalsandonlinetestingpartii/

Rothman,R. & Marion, S.F (2016). The next generation of state assessment and accountability. *Phi Delta Kappan, 97*(8), 34-37.

Schaeffer, B. (2012, August). *Resistance to High Stakes Testing Spreads.* Retrieved from https://www.fairest.org/resistance-high-stakes-testing-spreads

Schaffhauser, D. (2011). High stakes online testing: Coming soon! *THE Journal*, *38*(6), 28, 32, 34, 36. 38-39.

Schauffhauser, D. (2013, July). Is your school tech ready for Common Core? *THE Journal*. Retrieved from https://thejournal.com/Articles/2013/07/10/Getting-Your-School-Tech-Ready-for-Common-Core-Assessments_0.aspx

Schoenfeld, A. H. (2015). Summative and formative assessments in mathematics supporting the goals of the Common Core Standards. *Theory Into Practice, 54*, 183–194. doi:10.1080/00405841.2015.1044346

Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened test anxiety among young children: Elementary school students' anxious responses to high-stakes testing. *Psychology in the Schools, Vol. 50,* 489-499. doi:10.1002/pits.21689

Sharp, L. A. (2016). ESEA reauthorization: An overview of the Every Student Succeeds Act. *Texas Journal of Literacy Education, 4*(1), 9-13.

Smith, J.M., & Kovacs, P.E. (2011). The impact of standards-based reform on teachers: The case of 'No Child Left Behind'. Teachers and Teaching: Theory and Practice, 17(2), 201-225. Retrieved from http://dx.doi.org/10.1080/13540602.2011.539802

Strain-Seymour, E., Craft, J., Laughlin Davis, L., & Elbom, J. (2013). *Testing on tablets: Part I of a series of usability studies on the use of tablets for K-12 assessment programs* (White Paper). Pearson. Retrieved from https://www.pearson.com/efficacy-and-research/schools-education-research/research-reports/assessment/testing-on-tablets.html

1(1), 14-23.

Walker, T., (2015, December). *neaToday, 15*(12), 09. Retrieved from http://neatoday.org/2015/12/09/every-student-succeeds-act/

APPENDIX A

Consent Letter

Georgia Southern University
Office of Research Services & Sponsored Programs

Institutional Review Board (IRB)

Phone: 912-478-5465
Fax: 912-478-0719
IRB@GeorgiaSouthern.edu
Veazey Hall 3000
PO Box 8005
Statesboro, GA 30460

To: Brown, Doriastion
Chamblee, Gregory

From: Office of Research Services and Sponsored Programs

Initial Approval Date: 5/23/2017

Expiration Date: 4/30/2018

Subject: Status of Application for Approval to Utilize Human Subjects in Research – **Expedited Process**

After a review of your proposed research project numbered **H17369** and titled **"Grades 3-5 Teachers' and Principals' Perceptions Regarding On-Line Georgia Milestones Testing: A Quantitative Study of Rural Georgia School District."** it appears that (1) the research subjects are at minimal risk, (2) appropriate safeguards are planned, and (3) the research activities involve only procedures which are allowable. You are authorized to enroll up to a maximum of __68__ subjects.

Therefore, as authorized in the Federal Policy for the Protection of Human Subjects, I am pleased to notify you that the Institutional Review Board has approved your proposed research. **Description: The purpose of this study is to determine perceptions of principals and teachers regarding high stakes testing, specifically the Georgia Milestone Assessment System.**

If at the end of this approval period there have been no changes to the research protocol; you may request an extension of the approval period. In the interim, please provide the IRB with any information concerning any significant adverse event, **whether or not it is believed to be related to the study**, within five working days of the event. In addition, if a change or modification of the approved methodology becomes necessary, you must notify the IRB Coordinator **prior** to initiating any such changes or modifications. At that time, an amended application for IRB approval may be submitted. Upon completion of your data collection, you are required to complete a *Research Study Termination* form to notify the IRB Coordinator, so your file may be closed.

Sincerely,

Ele Haynes (SD

Eleanor Haynes
Compliance Officer

APPENDIX B

Permission To Use Survey

RE: Request to use survey

[redacted]

James L Pate <jlpate@valdosta.edu>

Reply

Fri 10/14/2016, 10:04 AM

You have our permission to use the survey requested.

Leon Pate

APPENDIX C

SURVEY

Please rate each statement below in terms of perception.

(1) Strongly Agree (2) Agree (3) Undecided (4) Disagree (5) Strongly disagree

_____1. On-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach.

_____2. On-line high stakes testing is counter to the idea of a balanced curriculum (equal attention to subjects).

_____ 3. On-line high stakes test items accurately reflect the content students learn through a school's curriculum.

_____ 4. Students' scores on on-line high-stake tests provide feedback for schools to improve the curriculum.

_____ 5. On-line high stakes test content is aligned with the school's curriculum.

_____ 6. Students' scores on on-line high stakes test are a valid measure of teaching ability.

_____ 7. The quality of teachers' instruction is directly related to student performance on on-line high stakes test.

_____ 8. On-line high stakes testing requires test preparation that diminishes time to teach subject content.

_____ 9. On-line high stakes testing has increased cooperation among teachers.

_____10. On-line high stakes testing has increased teacher and principal cooperation.

_____11. On-line high stakes testing increases when the school receives a failing grade.

_____12. Teachers experience stress in the effort to maintain their school's accountability grade.

_____ 13. District supervisors' pressure to improve on-line high stakes test scores increases teacher stress.

_____14. Principals' pressure to improve on-line high stakes test scores increases teacher stress.

_____15. Punitive measures associated with on-line high stakes testing increase teachers stress.

_____16. On-line high stakes testing has increased teachers' awareness of the accountability issues in education.

_____17. On-line high stakes testing is a reform measure that improves the quality of education.

_____18. Teachers are more accountable because of on-line high stakes testing.

.

Please briefly answer the following four questions:

What is your opinion of on-line testing?

What is your opinion of the accuracy of on-line high stakes testing?

What are barriers you see to on-line high stakes testing?

How do you feel on-line high stakes testing has changed instruction?

Other comments:

APPENDIX D

Administrators Survey Results (N=8)

Item #	Question	School A (N = 4)		School B (N = 4)		Combined (N = 8)	
		M	SD	M	M	SD	M
1	On-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach.	2.0	0.0	1.5	0.57	1.75	0.46
2	On-line high stakes testing is counter to the idea of a balanced curriculum (equal attention to subjects).	1.5	0.57	2.0	0.0	1.75	0.46
3	On-line high stakes test items accurately reflect the content students learn through a school's curriculum.	1.0	0.0	1.0	0.0	1.0	0.0
4	Students' scores on on-line high-stake tests provide feedback for schools to improve the curriculum.	1.0	0.0	1.0	0.0	1.0	0.0
5	On-line high stakes test content is aligned with the school's curriculum.	4.25	0.57	4.0	0.0	4.13	0.35
6	Students' scores on on-line high stakes test are a valid measure of teaching ability.	2.0	0.0	4.25	0.5	4.37	0.52
7	The quality of teachers' instruction is directly related to student performance on a high stakes test.	4.5	0.57	1.7	0.5	1.87	0.35
8	The quality of teachers' instruction is directly related to student performance on a high stakes test.	3.75	0.5	4.25	0.5	3.75	0.46
9	On-line high stakes testing has increased cooperation among teachers.	1.0	0.0	3.75	0.5	3.75	0.46
10	On-line high stakes testing has increased teacher and principal cooperation.	2.75	1.25	1.75	0.5	1.38	0.52
11	On-line high stakes testing increases when the school receives a failing grade.	3.0	0.8	3.0	0.81	3.0	0.76
12	Teachers experience stress in the effort to maintain their school's accountability grade.	1.75	0.5	2.0	0.0	1.87	0.35
13	District supervisors' pressure to improve high stakes test scores increases teacher stress.	2.0	0.0	2.0	0.0	2.0	0.0
14	Principals' pressure to improve high stakes test scores increases teacher stress.	2.5	0.5	2.0	0.0	1.87	0.35
15	Punitive measures associated with high stakes testing increase teachers stress.	4.25	0.5	4.0	0.81	4.13	0.64
16	High stakes testing has increased teachers' awareness of the accountability issues in education.	1.0	0.0	1.0	0.0	1.0	0.0
17	High stakes testing is a reform measure that improves the quality of education.	4.0	0.0	4.0	0.0	4.0	0.0
18	Teachers are more accountable because of on-line high stakes testing.	4.0	0.0	4.0	0.0	4.0	0.0

APPENDIX E

Teachers Survey Results (N=60)

Item #	Question	School 1		School 2		Both	
		M	SD	M	SD	M	SD
1	On-line high stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach.	1.22	0.42	1.06	0.25	1.18	0.39
2	On-line high stakes testing is counter to the idea of a balanced curriculum (equal attention to subjects).	1.11	0.32	1.33	0.48	1.16	0.37
3	On-line high stakes test items accurately reflect the content students learn through a school's curriculum.	3.11	0.32	2.33	0.48	3	0.71
4	Students' scores on on-line high-stake tests provide feedback for schools to improve the curriculum.	3.68	0.8	3.6	0.73	3.66	0.7
5	On-line high stakes test content is aligned with the school's curriculum.	1.84	0.36	1.8	0.41	1.83	0.37
6	Students' scores on on-line high stakes test are a valid measure of teaching ability.	2.64	2	3.46	0.74	3.53	0.74
7	The quality of teachers' instruction is directly related to student performance on a high stakes test.	4.15	4.5	4.13	0.35	4.15	0.36
8	The quality of teachers' instruction is directly related to student performance on a high stakes test.	2.64	3.75	2.3	0.51	2.56	0.62
9	On-line high stakes testing has increased cooperation among teachers.	2.17	1	3.06	1.27	2.4	1.12
10	On-line high stakes testing has increased teacher and principal cooperation.	3.11	2.75	3.8	0.56	3.28	0.76
11	On-line high stakes testing increases when the school receives a failing grade.	3.04	0.63	3.06	0.25	3.05	0.56
12	Teachers experience stress in the effort to maintain their school's accountability grade.	1.17	0.38	1.2	0.41	1.18	0.39
13	District supervisors' pressure to improve high stakes test scores increases teacher stress.	1.02	0.14	1.06	0.25	1.03	0.18
14	Principals' pressure to improve high stakes test scores increases teacher stress.	1.35	0.48	1.13	0.35	1.3	0.46
15	Punitive measures associated with high stakes testing increase teachers stress.	1.55	0.75	1.86	0.51	1.63	0.71
16	High stakes testing has increased teachers' awareness of the accountability issues in education.	2.4	0.53	1.13	0.35	2.08	0.74
17	High stakes testing is a reform measure that improves the quality of education.	4.68	0.46	4.86	0.35	4.73	0.44
18	Teachers are more accountable because of on-line high stakes testing.	4.0	0.63	4.13	0.35	4.03	0.58