January 2013

The Implications Of Accountability Models Weighing The Same Student Data Differently

Neely Traylor
Eastern Kentucky University

THE IMPLICATIONS OF ACCOUNTABILITY MODELS WEIGHING THE SAME STUDENT DATA DIFFERENTLY

By Neely R. Traylor

Dissertation Approved:

Chair, Advisory Committee

Member, Advisory Committee

Member, Advisory Committee

Member, Advisory Committee

Dean, Graduate School

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a Ed. D degree at Eastern Kentucky University, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of the source is made.

Permission for extensive quotation from or reproduction of this dissertation may be granted by my major professor, or in his absence, by the Head of the Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature Heidi R. Traylor

Date 11/30/13

IMPLICATIONS OF ACCOUNTABILITY MODELS

WEIGHING THE SAME STUDENT DATA DIFFERENTLY

By

NEELY R. TRAYLOR

Master of Arts
Eastern Kentucky University
Richmond, Kentucky
2008

Masters of Arts
Western Kentucky University
Bowling Green, Kentucky
2002

Bachelor of Arts
Lindsey Wilson College
Columbia, Kentucky
1999

Submitted to the Faculty of the Graduate School of
Eastern Kentucky University
in partial fulfillment of the requirements
for the degree of
Doctorate of Education
December, 2013

DEDICATION

This dissertation is dedicated to my husband, Chad, and daughters, Abby Rose and Elly Jane. Thank you for your patience as I missed many family dinners, bedtime stories, homework sessions, and even a few family outings to complete this process. Although I will never get those hours back, I pray I have instilled the importance of education and finishing what you start. Chad, I could have never completed this without you and will forever be in debt to you for allowing me to fulfill this dream.

ACKNOWLEDGMENTS

I would like to begin by thanking God for giving me the patience and willpower to complete this process. He is always beside me, blessing me with more than I deserve. To my dissertation chair, Dr. Hausman, thank you for your guidance and career advice along the way. Dr. Bliss and Dr. Biggin, thank you for your patience and answering my never-ending questions. To Dr. Lori Hollen, words cannot express my gratitude for your help in this process and in life; you are a wonderful friend and colleague. To my parents, thank you for instilling in me the determination to stay the course. That lesson alone has taken me far in life. Finally, a sincere thanks to every educator along my journey who encouraged and inspired me to stay the course. There are dozens; however, two stand out in my mind: Mrs. Patty Evans and the late Mr. Larry Williams. May every student have at least one teacher who believes in and supports them unconditionally.

ABSTRACT

There is little debate about whether or not the U.S. education system is in need of reform. Some people believe that public education reform will come about with high-stakes accountability and policies like *No Child Left Behind* (2001). States around the nation are facing similar questions. For example, how do we best determine college and career readiness and assess the quality of the education provided by our schools? The most common indicator of quality in public schools has been achievement test scores. This study examined three elementary and two high school drafts of the Kentucky educational accountability system known as the Unbridled Learning Accountability Model, which was adopted by the Kentucky Board of Education as a result of the *No Child Left Behind Act*. In addition, the study explored the implications of each of these drafts on schools and districts throughout the commonwealth using the same student data. The analyses of these data generated hypotheses about which kinds of schools might have been negatively and positively affected by each of the five accountability models that were under consideration.

Table of Contents

List of Tables

CHAPTER ONE

INTRODUCTION

Overview

This investigator examined three elementary and two high school drafts of the Kentucky educational accountability system known as the Unbridled Learning Accountability Model, which was adopted by the Kentucky Board of Education as a result of the *No Child Left Behind Act.* In addition, the study explored the implications of each of the drafts on schools and districts across the Commonwealth using the same set of student data. The analyses of these data provide information about which types of schools are negatively and positively affected by each of the five models that were under consideration.

Background

As U.S. public school policy moves from providing a free public education to all, first proposed by Thomas Jefferson in Virginia, toward equal education for all promoted by Horace Mann and W.E.B. Dubois, education practitioners and policymakers alike have struggled with how to best determine the quality of public elementary and secondary schools (Meier, Kohn, Darling-Hammond, Sizer & Wood, 2004). It has become evident that preparing young students to be "competitive in a global information age requiring a skilled workforce," and our "future economic success" will require the preparation of a larger percentage of post-secondary graduates than ever before (Maruyama, 2012, p. 252). Reardon (2011) argued that "nations with a more skilled

population grow faster than those with a less skilled population" (p.3). States around the nation are facing the same questions; How do we best determine college and career readiness and assess the quality of the education provided by our schools? The most common indicator states are currently using to measure the quality of education is our students' achievement test scores. However, the assessments and accountability models being used to determine school and even teacher quality are, at best, causing confusion for educators and the public alike. One source of confusion are the public policies on labeling schools. A school may be labeled as successful according to one accountability system but needs improvement by another system (Linn, 2006; Meier et al., 2004).

There has been little debate that our public education system could better prepare our students to compete globally and improve our economy. Meier et al. (2004) explain "America's economic, social, and moral strength still depends on it [education]" (p. xvii). There are stakeholders who believe this much needed reform will come about with high-stakes accountability and policies like *No Child Left Behind* (*NCLB*, 2001). The idea of using high-stakes assessments and accountability systems like *NCLB* to hold educators and students accountable in education is certainly not a new one; however, stakeholders are also concerned with astronomical monetary and human costs associated with these assessments and accountability models (Wiliam, 2010). Educators, policymakers, test creators, and other stakeholders around the nation spend countless hours and billions of dollars to design assessments and accountability models, in hopes the assessments will effectively and efficiently assess student learning and provide data to be analyzed and used to improve teacher instruction, student achievement, and hold students, teachers,

and schools accountable (Marcus, 2012; Meier et al., 2004; Wiliam, 2010;). These paradigm shifts are bringing about major changes for America's schools.

One major change in educational accountability policies is the shift from measuring inputs (e.g. expenditures per pupil, quality of teachers, curriculum, policies and practices) to measuring outcomes (e.g., student achievement scores) (Ananda & Rabinowitz, 2001; Choi, Goldschmidt, & Yamashiro, 2006). However, as many studies have illustrated, there are several concerns with high-stakes assessment and accountability models used to measure the outcomes (Meier et al., 2004; Popham, 2001; Popham, 2003). One concern critics of high-stakes assessments have is that more often than not, these tests are used to determine factors they were not designed to assess, thus affecting the validity and reliability of the accountability model (Popham, 2001). Another concern with American educational accountability, in contrast to accountability models used around the world, is that the stakes are much higher for teachers than students (Wiliam, 2010). Critics also cite research reporting that the same test data can supply glaringly different results depending upon the accountability model applied to the data, holding different schools accountable (Yu, Kennedy & Teddlie, 2007). Finally, many educators also struggle with a single accountability measure and contend that "standardized tests cannot adequately or accurately measure school performance…" calling for a multiple measure approach (Brookhart, 2009; Chester, 2005; Choi et al., 2005, p.2; Schafer, 2003).

With the revisions of the *Elementary and Secondary Act* (ESEA) and *No Child Left Behind* (*NCLB*), an increased focus on accountability and consequences emerged. The measure used by *NCLB* to operationalize accountability is Adequate Yearly Progress

(AYP), which focuses on achievement producing a percent proficient indicator. Criticisms of the AYP measure include the fact that in many states, AYP does not consider growth, only student proficiency at one point in time, and identifies too many schools as failing. As Perie and Park (2007) explained, there are several ways to miss AYP and only one way to meet it; a "school must meet its target for every single subgroup to meet AYP, but if it misses just one—no matter which one or how many over one—it is placed into an improvement category" and "schools serving more diverse students were less likely to meet AYP requirements than schools serving less diverse students" (p. 13).

Current research surrounding the selection of the best accountability model for determining school performance and identifying low-performing schools suggests that value added models (VAMS) are the most valid (Amrein-Beardsley, 2008; Betebenner &Linn, 2009; Ready, 2012; Yu et al., 2007). Further, in the last two decades, there has also been a call for the integration of student growth into educational accountability models (Choi et al., 2005; Linn, 2005 B; McEachlin & Polikoff, 2012; Yu et al., 2007). However, there is still a great deal of research needed to determine the appropriate uses of accountability models, specifically the value-added models and student growth models being used by many states to measure teacher effectiveness.

The above studies clearly offer research on determining the best accountability models for specific purposes. Further, they offer state education departments essential information for use in deciding upon the specific accountability model to be used within an accountability system. However, for the interest of this study, these studies do not research the implications of the revisions to the same accountability system using the

same dataset. Researching these implications will aide educational leaders and policymakers in determining the best accountability model for their needs.

Statement of the Problem

This study addressed a perceived need in the literature for research on the implications of different versions of the same accountability system weighing the same student data differently. Ample research has been completed to determine the variances among different accountability models such as: value-added models, growth models, status models, and improvement models. There is, however, a specific need to determine if different versions of the same accountability systems using the same data will provide like results or provide results that differ significantly. Furthermore, if the results differ, policymakers need to know how these various models impact schools with different characteristics differently. One concern for schools and districts across the state is the identification of the highest and lowest performing schools in the state. Because of the rewards, recognition, and consequences associated with such labels, there is a need to be very intentional in determining what a state accountability model will be designed to measure and choosing the model that most reliably utilizes that data. Creating an accountability system that provides data that best identifies a school or district's performance on the specific variables defined and consistently identifies the same high and low performers is crucial to the reliability of high stakes accountability models and systems.

The Purpose Statement

The primary purpose of this study was to explore the implications of five different drafts, three elementary and two high schools, of Kentucky's Unbridled Learning Accountability Model using the same student data. Another purpose of this study was to inform school, district, state, and national educators; policymakers; and other stakeholders to help ensure they clearly determine what variables will be measured by their accountability systems and create the most effective assessments and accountability models possible to measure these determined variables. Further, this study provides research for the stakeholders to make them more aware of the implications of their work for districts and schools of differing demographics. This study also provides a rationale for being very intentional in the steps in creating an accountability system: first, be very intentional in determining what exactly will be measured; second, identify which accountability model is most valid and reliable in measuring these variables; and third, be intentional and careful designing an accountability system that accurately and consistently measures school and district performance and consistently identifies the high and low performing schools and districts on the determined variables. Lastly, those designing accountability models must also determine how the results will be transparent and communicated to all stakeholders.

Research Questions

1. To what extent would changes to Kentucky's Unbridled Learning Accountability Model result in different conclusions when utilizing the same student data?

2. Which school types are positively/negatively affected by each of the Kentucky models that were under consideration circa 2010?

Research question number one addresses changes in a schools reward or assistance categories schools could be placed in based on their accountability scores. Specifically, the research studied the top and bottom five percent of elementary and high schools based on their total accountability score. Research question number two sought to determine if schools with specific characteristics were positively or negatively impacted. These characteristics include school size, proportion of non-white students, proportion of students eligible for free and reduced lunch rates, per-pupil expenditure, average years of teaching experience, and teacher education.

Rationale for Study

There are several studies addressing the validity, reliability, and other concerns with each of the different accountability models most widely used by states to assess student achievement, with each study offering suggestions for choosing a model (Betebenner & Linn, 2009; Choi et al., 2006; Lauermann &Karabeinck, 2011; Yu et al., 2007), but these studies do not address how variations of the same accountability model would impact schools and districts. Another study similar to this one addressed which schools would be identified as lowest performing under a proposed revision of the *Elementary and Secondary Education Act* (McEachlin & Polikoff, 2012). However, that study used a more global approach to and looked at the implications of national policies that impact the development of state accountability systems, whereas this study focuses

on the implications of one accountability system and the drafts leading to the current model upon the schools and districts of Kentucky.

Significance of Study

This study is significant because of the importance placed upon the success of current educational reform in America. The United States educational system is no longer ranked among the top internationally. Kress, Zechmann and Schmitten suggested "a growing and widespread concern about the vital importance of education to our national security" (2011, p. 188). If we are to succeed in reclaiming our place among the top education systems in the world, we must accurately assess the success of our system. The success of our nation's educational system is primarily measured by high-stakes accountability tests, models, and systems. The study of the implications of different high-stakes accountability models using the same student data is imperative to current education reform because we must ensure our educational systems are identifying the schools and districts that are best producing students with the 21st century skills needed to compete globally and the economic success of the United States. Further, we must be able to accurately identify those schools and districts not producing students with these skills because research strongly suggests these students will not only most likely be required to take remedial courses if they pursue post-secondary education, but they will earn significantly less income over the course of their lifetime. Table 1.1 illustrates the discrepancy in income based upon education. The Hamilton Project reports that nearly 80% of high school dropouts made less than $30,000 in 2010, while 80% of college graduates earned around $100,000 (Greenhouse, Harris, Karen, Looney, & Patashnik,

2012). Further, The McKinsey Global Institute suggests there will be a global shortage of $38 to $40 million college educated workers by 2020. Further, the group predicts there will be a potential shortage of 44 million workers with the secondary education needed to be qualified for labor-intensive manufacturing and services (Dobb, R., Madgavkar, A., Burton, D., Labaye, E., Manyika, J., Roxburgh, C., Lund, S., & Madhav, S. 2012).

Table 1.1
Income over Course of Lifetime Based on Education

Education Level	Average Lifetime Earnings
Professional Degree	$4.4 Million
Doctoral Degree	$3.4 Million
Master's Degree	$2.5 Million
Bachelors Degree	$2.1 Million
Associate's Degree	$1.6 Million
Some College	$1.5 Million
High School Graduate	$1.2 Million
Non-High School Graduate	$1 Million

Source: Greenhouse, M., Harris, M., Li, K., Looney, A. & Patashnik, J. (2012). A dozen economic facts about k-12 education. The Hamilton Project, Washington, D.C.

Limitations of Study

Limitations to this study are varied in nature. A few limitations revolve around the data available from Kentucky's Unbridled Learning Accountability Model. For example, the 2011-2012 school year was the first administration of the K-PREP testing system used for the accountability model. Because this was the first administration, stakeholders were unsure of the alignment between the standards and the developed tests. Another limitation for the Kentucky Department of Education and school districts across the state is the limited resources at this time of economic stress. Because of *NCLB* and Senate Bill 1, the Kentucky Department of Education had to develop an entirely new testing system and accountability model. School districts had very limited resources and time to take

the steps necessary to quickly align curricula to a new set of standards and prepare for the new tests they knew very little about during the first administration. In addition, this study is based upon the current model adopted by the Kentucky Board of Education; however, the model is evolving with new data and research. Lastly, the results of this study are limited to the study of the accountability model for the state of Kentucky only. Other states may wish to conduct the same type of study on their accountability model.

CHAPTER TWO

LITERATURE REVIEW

Overview

This study addressed the implications of different revisions of the same accountability system using the same student data. As this literature review will illustrate, there is ample research to determine the strengths and limitations among different accountability models such as: value-added models, growth models, status models, and improvement models, as well as the differing effects of these models. There is, however, a specific need to determine if different versions of the same accountability systems using the same data provide similar results or different results. Further, this study determined which types schools are positively and negatively affected by each of the drafts.

This study also explored the implications of five different drafts of Kentucky's Unbridled Learning Accountability Model utilizing the same student data. The study involved three elementary and two high school drafts proposed for the Unbridled Learning Accountability Model. The results may be used to inform school, district, state, and national educators; policymakers; and other stakeholders to help ensure they clearly determine what variables will be measured by their accountability systems and create the most effective assessments and accountability models possible to measure these determined variables. It also provides research for the stakeholders to make them more aware of the implications of their work on districts and schools of differing demographics. Finally, this study provides a rationale for being very intentional in the

research-based steps in creating, implementing, and reporting the results of an effective accountability system.

This chapter is organized around several topics: a brief history and rise of accountability in the United States; the case for accountability; the case against accountability; federal accountability; state accountability; explanation of the most popular accountability models; accountability model concerns; and research on variations in accountability using different models.

History and Rise of Educational Accountability

Accountability is not a new idea in the field of education. Taxpayers, parents, employers, and other stakeholders view accountability as a tool to determine if "…instruction has had its intended effect" (Wiliam, 2010, p. 107) and as a way to gauge our success in preparing our students to "…meet the challenges of a globalized, technology-driven economy" (Greenhouse et al., 2012, p. 9). Concerns and criticisms regarding the introduction of "high-stakes" testing began in the United States as early as the 1830s (Wiliam, 2010). Following the success of the United States Army's Alpha test, designed by Alfred Binet, during World War I, there was a fluctuation of tests being designed to make comparisons among test-takers by comparing their performances to that of a norm group. These tests were "intelligence focused aptitude tests" and "achievement tests" (Popham, 2001, p. 42).

Well into the 1960s, Americans seemed pleased with public education and the assessments being used. However, discontent with public education began to build during the 1970s with newspapers publishing articles about students who could not read but

were promoted because of "seat time" (Popham, 2001, p. 4). Because of stories like these and the "citizen distress", national, state, and local legislators; education officials; and school boards began implementing minimum competency tests for grade promotion and/or graduation (Popham, 2001, p. 4; Kress, Zechman, & Schmitten, 2011). Linn (2005) reports that "two-thirds of the states introduced some form of minimum competency testing during the 1970s and early 1980s" (p. 2). The "minimum competency tests are intended to return meaning to the high school diploma by requiring that students meet various indicia of basic competence" (Kress et al., 2011, p. 190; Shephard, 1980). Since the minimum competency tests of the 1970s and 80s, educational accountability has become increasingly important because elected officials and policymakers have limited ways to impact instruction or gauge the effectiveness of education. Accountability tests are used as the "policy tools to hold teachers and school administrators accountable for student learning" (Linn, 2006, p. 3).

The Case for Accountability

Over time, educational accountability has grown from systems designed to evaluate schools and allocate sanctions and rewards, like Kentucky's previous accountability system, the Commonwealth Accountability Testing System (C.A.T.S.), to "… a system that allows the public to understand how well their schools are working and to provide information to policymakers on the changes that are needed to make the schools more effective and to continually improve all students' educational opportunities" (Perie, Park, & Klau, 2007, p. 4). According to Robert Linn (2000), distinguished professor for the Center for Research on Evaluation, Standards, and Student

Testing, tests and assessments have been used for over 50 years to track and select students for special programs like gifted and talented, special education, and English language development; they have also been used for program accountability, minimum competency testing, school and district accountability. These assessments are at the core of the standards based accountability systems of the 1990s to present. Policymakers tend to act on the belief that assessments are particularly effective change agents because they are inexpensive in comparison to other reforms like increasing instructional time and reducing class size. Testing and assessment also can be externally mandated and rapidly implemented. Perhaps most importantly, the results are visible. Stakeholders are able to see black and white results to determine the effectiveness of public education (Linn, 2000). Perie and Park (2007) explain that the communication about an accountability system's goals and consequences during its inception and the understandable and useful reporting of results are both very important to the effectiveness of an accountability system. These results and the feedback provided by an effective accountability system can often improve performance without additional interventions (Baker, 2005; Tankersley, 2007).

Haertel and Herman (2005) noted that uses of tests and accountability models have varied over the years; however, there are several positive uses of the data:

> … help clarify expectations for teaching and learning; monitor educational progress of schools and students; monitor the progress of demographic subgroups of students and gaps in achievement of those subgroups; encourage the closing of the gaps in performance among racial/ethnic subgroups and between economically disadvantaged students and their more affluent peers; motivate greater effort on the part of students, teachers, and school administrators; contribute to the evaluation of educational programs and schools; identify schools and programs that need to be improved; and provide a basis for the distribution for rewards and sanctions to schools and students (p. 3).

With the national spotlight on America's education system, stakeholders desire an effective and transparent measure of student achievement, instruction, and in many cases, teacher quality. Although there is tremendous controversy and debate about some of the measures used to gauge these variables, most stakeholders agree there must be accountability; deciding on what is to be measured and how it will be measured, however, is not something on which stakeholders can easily agree.

The Case Against Accountability

Although the *No Child Left Behind Act* (2001) was not the origin of high-stakes assessment and teacher accountability, this policy has brought assessment and accountability to the forefront in education reform. Choi et al. (2006) contend that "In order to hold schools accountable for student performance, there is an implicit assumption that it is possible to isolate a schools' effect from all other factors that might influence achievement (e.g., student background or inputs outside of a school's control)" (p. 4). Many researchers believe schools should only be held accountable for the factors or variables they can control (Bathgate, Colvin & Silva, 2011; McEachin & Polikoff, 2012; Yu et al., 2007). Despite these concerns, numerous states have adopted accountability models that are based on the idea that student achievement scores are an accurate indicator of school and teacher effectiveness in delivering effective instruction. Further, these assessments and accountability models often produce results that do not exclude variables outside of the school's control (Lauermann &Karabenick, 2011).

According to Popham (2001), the most important consideration when creating and using high-stakes assessment is how the test will be used to improve the instruction

students receive. Jones and Egley (2007) argued, "High-stakes testing may never be able to provide teachers with the type and level of feedback they need to improve their instruction" (p. 246) because it is so often misused. Criticisms like this lead many to question the focus upon test results and teacher accountability. However, political leaders, community leaders, and taxpayers searched for a way to ensure an increase in student achievement, and the pressure that policies like *NCLB* bring is believed to be the cure all for our education woes. With just enough pressure and consequences, but often times a severe lack of funding, policymakers and leaders expect educators to do what many believe is unrealistic- bring every student to proficiency in reading and mathematics by 2014.

The palpable feeling of pressure to achieve such a daunting and unrealistic goal is a constant source of tension for American educators, leaving many teachers unhappy in their choice of profession. Jones and Egley (2007) noted that 90% of North Carolina teachers reported that their jobs were more stressful since the implementation of *No Child Left Behind* because of the numerous mandated tests and accountability. This pressure reported by our nation's teachers also leads teachers to instructional practices that do not align with their personal values. For example, Quinn and Ethridge (2006) reported that children in many public schools are spending a large portion of their school day being taught how to take standardized tests, and they are studying only what is thought to be on the test. Although many teachers know these instructional practices are not "best-practice," the pressure to perform well on high stakes assessments leads them to these practices, leaving educators feeling frustrated and powerless to do what they feel is best for students.

Education critics have argued that educational policies like *NCLB* are inconsistent with findings of research and years of professional practice (Scot, Callahan & Urquhart, 2009). Teachers feel like their efforts to use research based "best practices" are not valued, and they resent the loss of their autonomy (Finnigan & Gross, 2007) brought about by *NCLB* and the standards movement. Rather than acknowledging teachers for the professionals they are, *NCLB* is pushing schools into implementing drastic curricula changes that are intended only to help students achieve on mandated assessments (Herrera & Murry, 2006; Ho, 2008; Lee, 2008). Unfortunately, this pressure to perform has many negative implications for teachers and students.

One implication of high stakes accountability is the fact that many teachers report they are pressured to teach to the test. Popham (2001) explained that if constructed properly and used correctly, educational tests can help educators improve their instruction. However, the misdirected pressure upon teachers forces them to teach students how to do well on tests rather than teaching each student what they need to know (Popham, 2001). In a study of Florida teachers (Jones & Begley, 2007), 43% of mathematical and writing instruction time per week was spent on test taking-strategies. Similarly, 38% of reading instructional time was spent on test-taking strategies (2007). Jones and Begley (2007) also highlighted a point that research emphasizes the importance of teaching for understanding. However, as teachers feel pressured to teach to the test, students are not learning for understanding; they are learning to recall information. This level of learning leaves students without the ability to analyze, synthesize, or justify their thoughts; in the end, students are not achieving increased levels of learning, the major goal of *NCLB* (Lee, 2008; Meier et al., 2004).

Not only are students spending significant amounts of time learning test-taking strategies, they are, in large, only learning what educators feel will be tested. Research indicates that due to the pressures to score well on achievement tests, educators are narrowing the curriculum to what is tested, with little to no time spent on non-tested areas like music, art, and physical education (Herman & Dietel, 2005; Popham, 2001). Subjects like science and social studies also are often reduced or not taught if they are not tested. Lee (2008) contends that the "inflated test scores" which often result from intensive drills and curricular reduction "give the false impression" that interventions and instruction are working (p. 611).

Amrein-Beardsley, Berliner, and Rideau (2010) posit that if teachers know their performance will be based upon their students' performance on test results, they are given an incentive to cheat. Because of *NCLB* and the pressures associated with the policy, more teachers and administrators are cheating. In 2010, *New York Times* Columnist Trip Gabriel (2011) reported that one in five elementary and middle schools in Georgia submitted score reports that were considered abnormal with 90% of one school's reports labeled as suspect. Consequently, the Georgia State School Board ordered the investigation of 191 schools based upon 2009 reading and math tests results. Martel (2011) explained that this investigation involved 178 principals, teachers, or other staff and concluded in an 800 page report that lead to the referral of 11 teachers and administrators to the state agency with the power to revoke their licenses. John Fremer, a data forensics specialist, indicated, "Every time you increase the stakes associated with any testing program, you get more of cheating" (Gabriel, 2010). The stakes are certainly increasing; as Gabriel points out, Colorado passed a law making a teacher's tenure

dependent upon student scores. Almost a dozen states, including Kentucky, plan to evaluate teachers based on some form of student data or test scores, and some are offering bonuses to teachers based on student achievement scores. Jones and Egley (2007) argued that primarily focusing upon test scores can have negative implications that prove to be "detrimental" to students' education. They further emphasize that high-stakes data are better suited for assessing district or school-wide trends and should not be used alone to determine student learning or teacher effectiveness.

The misuse of data and low level of confidence in the validity of test data to actually gauge student achievement are yet other concerns of educators, students, and parents. Choi et al. (2006) asserted that "it is crucial that valid inferences about school quality can be made" (p. 6); these valid inferences are commonly based on student scores on state-mandated assessments, and the test results are considered reliable when the test produces consistent results with each administration (2006). Popham (2001) suggested that our nation is taking part in a "one-size-fits-all" mentality with testing and argues that nationally standardized tests cannot accurately assess the effectiveness of all teachers because they are testing different students, using different standards, and "half of what is on the tests wasn't even supposed to be taught" (p. 43). An accountability model that is modeled on standards but uses an assessment that is not sufficiently aligned to the standards might incorrectly identify high or low performing schools because the standards and assessments are not sufficiently correlated (Choi et al., 2006; Rabibowitz, Roeber, Schroeder & Sheinker, 2006).

Many researchers are emphasizing that the most current accountability models utilize multiple measures to gauge student growth and achievement. Assessment experts

highly discourage any decisions being made upon one measure (Bathgate, Colvin & Silva, 2011; Chester, 2005: Lauermann &Karabenick, 2011; Meier et al., 2004; Schafer, 2003). However, educators and policymakers around the nation are going against research and misusing the data to make very important decisions like student placement and pay for teachers. Some are even misusing data to end employment of teachers and leaders.

Because of their lack of trust in the assessments and accountability systems being used to determine their effectiveness in achieving student growth and achievement, many teachers feel that high-stakes testing programs have stifled student learning by negatively impacting their teaching practices and forcing them to teach in ways that only promote test-taking skills rather than higher-order thinking (Jones & Begley, 2007). Jones and Begley recommend that state departments of education need to send clear messages on how test scores should be used (2007). Until the message is clearly sent to all stakeholders that test results are not the only measure that should be used to determine student achievement and growth, our students will continue to lose precious instructional time preparing for tests and taking these tests. Wiliam (2010) suggested that standardized tests are "inappropriate tools…to hold districts, schools, and teachers accountable" and that although there is a case for high-stakes accountability, "considerable work needs to be done to minimize the costs and maximize the benefits" (p.120). Linn (2006) noted that schools are receiving mixed messages between *NCLB* and their state accountability models. One of the systems, state or *NCLB*, may label the school effective, whereas the other model labels the same school or district as needing improvement. Issues like this cause confusion and skepticism for educators, parents, and the public.

Federal Accountability

A new era for public education dawned in the United States in 1965. Since the opening of the first American public school, Boston Latin School, in 1635, and Horace Mann's call for a "commonly educated public" (Kress et al., 2011, p. 187) during the 1890s, the *Elementary and Secondary Educational Act* (*ESEA*)- Title I- enacted in 1965 was the first congressional act involving the federal government in the functions of state and local school education affairs. *ESEA* was passed eleven years after the landmark *Brown versus Board of Education* case that ended segregation by race. *ESEA* was a "major piece of civil rights legislation due to its focus on improving the educational opportunities of poor children" (Kress et al., 2011, p. 189). Title I of *ESEA* not only provided financial support for those schools serving the nation's poorest children, but it also introduced testing requirements for Title I students (Linn, 2005 C).

The 1983 publication of *A Nation at Risk* by the U.S. Department of Education further increased the nation's attention to the need for education reform with a bleak description of the failings of the nation's educational system. Kress et al. (2011) state that the report played "upon the Cold War era fears, analogized its [education's] potential detrimental effects to that of a foreign act of war" (p. 190). The report argued that "…the nation's prosperity was imperiled and implied that other nations with better-educated populaces would overtake the U.S. economy if the education system were not reformed" (p. 190). *A Nation at Risk* also marked the beginning of a turning point in education testing and accountability, describing the need to shift from the minimum competency testing to input-focused reform, which called for more money, resources, and teachers. The report also planted the seed for the high expectations, proficiency standards, and

performance assessment movements requiring students to write extended or constructed responses, solve real-world mathematics problems, participate in performance events, and defend their solutions (Kress et al., 2011; Linn, 2006; Wiliam, 2010).

Those schools and districts heeding the report's suggestions "increased school budgets, decreased student/teacher ratios, increased credit requirements for graduation, increased the number of science classes students were required to take, lengthened school days and school years, and raised teacher salaries, among other measures" (Kress et al., 2011; Massell & Fuhrman, 1994). Wiliam (2010) found evidence that the authentic assessments (i.e. performance assessments and extended responses) have a significant positive impact on student learning. However, the high costs of such assessments and the missing reliability or technical quality caused most states to discontinue or greatly reduce their use of authentic assessments. Although the popularity of the authentic assessments waned, the standards movement remained strong. *Goals 2000: Educate America Act*, President Clinton's education initiative, reinforced the standards movement, calling for content standards, student performance standards, and the standards based approach to assessment and accountability (Linn, 2006).

A Nation at Risk also played an ironic role by increasing the federal government's role in state education. Two years prior to the publication of *A Nation at Risk*, President Ronald Reagan intended on closing the U.S. Department of Education. With the publication of this eye-opening report, the federal government's role in education policy became larger than ever and increased with each new administration.

According to Kress et al. (2011), the "consequential accountability movement itself began in the 1990s, when nearly forty states integrated the concept into their

education policies and Congress introduced it into federal education policies through the *Improving America's Schools Act*" (*IASA*) (p. 186). IASA, the predecessor of *NCLB*, mandated Title I evaluations, reinforced the standards-based approach to assessment and accountability, and allowed states to use high scores in one subject to compensate for low scores in another subject. Research on the consequential accountability systems of the 1990s indicated that significant increases in student achievement were made in those states that used them (Linn, 2006; Kim & Sunderman, 2005; Kress et al., 2011).

NCLB, which was signed into law in January of 2002 by President George W. Bush, made accountability "the centerpiece of the education agenda" (Wiliam, 2006, p. 1) by requiring states to develop and implement "consequential accountability" systems as a condition of receiving Title I federal education funds, which the federal government provides to schools and districts with a high percentage of students from low-income families (Kress et al., 2011, p.186). *NCLB* amended and reauthorized the *Elementary and Secondary Education Act*, *ESEA*, of 1965, requiring schools to test reading and mathematics every year in grades three through eight and once in high school; annual testing, evaluating school effectiveness, developing a timetable for student proficiency, and establishing sequential and specific consequences for failure were also requirements of *NCLB*. This historical educational legislation greatly expanded the role of the federal government in public education and reflected the views of "politicians, policymakers, and the business community" that the achievement among our students was lacking, reform was needed, and a lack of accountability attributed to the low performance in schools. In an effort to reform America's educational system, school accountability became the driving force of school reform. The thinking behind this force suggested that sanctions

and rewards would "prod" teacher and administrators to be more effective (Linn, 2006; Thum, 2003). *NCLB* was seen by many as "an evolution of previous attempts to use high-stakes tests to improve educational outcomes" (Wiliam, 2010, p.7).

NCLB has focused public attention on the performance of cohorts or groups of students, rather than individual students, making the proficiency rates more important than the growth of each student. *NCLB* requires that data for various groups be disaggregated, "including economically disadvantaged students, students with disabilities, students with Limited English proficiency, major racial and ethnic groups, and gender groups" (Kress et al., 2011, p. 214). This push to disaggregate data supported the mandate that all students must reach proficiency by 2014, thus placing emphasis on not only closing the gap between various groups but also raising the standards and achievement of all students at the same time.

The belief that the tests should measure the impact or quality of education provided by the school, not other factors like the amount of parental support was at the heart of *NCLB* (Wiliam, 2010). *NCLB* requires states to adopt "challenging academic content standards" which specify what students are expected to know and be able to do (Linn, 2006, pg. 4). The variability in the stringency of the state standards defining proficiency, however, is so great that the concept of proficient achievement lacks meaning (Linn, 2003 B). Another concern with the *No Child Left Behind Act* is expressed by Linn (2006) who contended, "It is clear that when combined with NCLBs multiple hurdles approach, disaggregation rules make it considerably more difficult for large schools with diverse students to meet … requirements than it is with homogenous student bodies" (Linn, 2006, p.14).

One of the causes of pressure upon educators, teachers, and students is a school's failure to meet Adequate Yearly Progress (AYP) as mandated by *NCLB*. Schools meet AYP requirements if the percentage of the students for the school as a whole and for each of the subgroups meet or exceed the Annual Measurable Objectives (AMO) in both reading/English language arts and mathematics (Linn, 2006; Yu et al., 2007). Schools and districts not meeting AYP for two or more consecutive years are held accountable and subjected to a series of punitive consequences, which vary from state to state. Hammond suggests that *NCLB* will label most of America's public schools as "failing" while they are actually improving student achievement. A study of California schools who did not meet AYP showed that they were designated as not meeting AYP not because their achievement levels were faltering but because a single student group, disabled students, Asian or African American students, or English as Second Language students had fallen short of the target. Schools with higher proportions of higher poverty and racially diverse students had a much higher chance of not meeting AYP (Kim & Sunderman, 2005; Linn, 2005A; Meier et al., 2004). This pressure alone leads schools to not include certain groups of students in these high-stakes tests. Many educators vehemently argue that English language learners and many students with disabilities should not be required to take the same tests as those with no disability and native English speakers. English language learners and some disabled students are at an obvious disadvantage on these tests (Fuller, Wright, Gesecki & Kang, 2007; Herman, 2007; Kim & Sunderman, 2005; Kress et al., 2011; Lee, 2008; Meier et al., 2004; Mintrop & Sunderman, 2009).

History of Educational Accountability in Kentucky

Kentucky's most contemporary accountability models include the Commonwealth Accountability Testing System (CATS), the result of the *Kentucky Education Reform Act of 1990* (KERA), and the Kentucky Performance Rating for Educational Progress (KPREP), Kentucky's response to the *No Child Left Behind Act*. In 1998, Kentucky's General Assembly enacted House Bill 53, which outlined guidelines for the replacement of the existing testing system and creation of a new testing system (CATS).

Commonwealth Accountability Testing System

Under the CATS accountability model, students took two tests, the Kentucky Core Content Test (KCCT), a criterion-referenced test, and the fifth edition of the Comprehensive Test of Basic Skills test (CTBS), a nationally norm referenced test (Seiler, Lunney, Olds & Young, 2005). CATS brought a new level of school accountability to Kentucky's public education system as a result of a landmark ruling from the state's highest court regarding the inequity of education resources.

On June 8, 1989, Kentucky's Supreme Court declared the state's public school system unconstitutional. In 1986, sixty-six Kentucky school districts filed a lawsuit, *Rose v. Council*, to receive "equitable and adequate" funding for public schools. The high court's ruling went far beyond the financing issues and ruled the entire educational system "deficient and unconstitutional." This landmark ruling made "certain student outcomes a constitutional obligation" and "held the state legislature responsible for monitoring the performance of the public schools," which "required an outcome based accountability system" (Foster, 1991, p. 34; Goldstein & Behuniak, 2005). Further, the

court required this new system be designed no later than April 15, 1990. The *Kentucky Education Reform Act* (*KERA*) of 1990, HB 940, was approved by the General Assembly and signed by Governor Wallace G. Wilkinson in April of 1990. The legislation became effective in June of 1990 and required increases in student achievement, attendance rates, and graduation rates. A large increase in state taxes funded a new education structure where schools meeting state goals received financial rewards, whereas those schools failing to meet these goals received technical assistance (Goldstein & Behuniak, 2005). Many found the Commonwealth Accountability Testing System superior to other models around the nation for several reasons; one reason being the CATS tested seven content areas, where many states tested fewer content areas, decreasing the chance of curricular reductionism (Linn, 2003A). Linn (2005B) also found the CATS accountability system comprehensive in nature. However, after twenty years of *KERA* and federal legislation, Kentucky overhauled its educational system again.

Kentucky's Unbridled Learning Accountability Model

The 2009 Kentucky General Assembly adopted Senate Bill 1, in response to the *No Child Left Behind Act*, which required Kentucky to revise the state's academic standards to be more rigorous. The legislature also mandated a new assessment and accountability system to include multiple measures of school and district effectiveness. The Unbridled Learning Assessment and Accountability System was first implemented in the 2011-2012 school year (Draut, 2011; Draut, K. & Sims, R., 2012). Kentucky was granted a waiver of the federal *NCLB Act* to become the first state to implement an "accountability model based on the goal of college and career readiness" (p.1) that also

addresses all of the aspects of school and district work, and is a more balanced approach to the Kentucky Board of Education's strategic priorities: next-generation learners, next-generation professionals, next-generation instructional programs and support, and next-generation schools and districts (Tungate, 2010). Dr. Terry Holliday, Kentucky Commissioner of Education, reports that the new system "is designed to provide in-depth information about the performance of students, schools, districts and the state as a whole" (2012, p.1). Rhonda Sims (2013), Kentucky Department of Education Office of Accountability and Assessment, explains that "Frequent, meaningful testing is required to assess the extent of student progress toward proficiency; accurate, understandable reporting is required so that all stakeholders in Kentucky education have the data needed for making effective decisions concerning school policies, programs and curricula" (p. 1).

The Unbridled Learning Accountability Model is comprised of five components that contribute points to the overall score for the Next-Generation Learners strategic priority. The five components are achievement, gap, growth, college and career readiness, and graduation rate (Gross, 2012; Kentucky Board of Education, 2010). Kentucky's new accountability model includes program reviews in non-tested areas like arts and humanities, practical living and career studies, and world languages. The model also places emphasis on the identification and closing of achievement gaps; providing support for the lowest performing schools; and "linking teacher and principal evaluation information to educator preparation programs," with the "ultimate goal" being to ensure all students are college and career ready (Gross, 2009, p. 31; Kentucky Board of Education, 2010).

Achievement Component of Unbridled Learning

The finalized achievement component of the Unbridled Learning Accountability system will be based on student performance on state administered tests in reading, mathematics, science, social studies, and writing on the Kentucky Performance Rating for Educational Progress (K-PREP) tests. High school students also take End-of-Course assessments in English II, Algebra II, Biology, and U.S. History. Student scores in these areas are labeled as novice, apprentice, proficient, or distinguished. For each content area, one point is awarded for each percent of students scoring proficient or distinguished. One-half point is awarded for each percent of students scoring apprentice, and no points are awarded for novice students. The goal of Unbridled Learning is 100% proficiency for all students.

Gap Component of Unbridled Learning

The Gap component of the Unbridled Learning Accountability system is determined by comparing the performance of students who are members of traditionally under-performing groups such as ethnic minorities, disability, low income, and limited English proficiency to the goal of 100% proficiency and to their peers who are traditionally higher performing. To calculate the combined student gap group score, non-duplicated counts of students who score proficient or higher and are in any of the student groups are added together, yielding a single gap number of proficient or higher students in the student gap group. One important difference between the Unbridled Learning Accountability Model and *NCLB* is that no student counts more than one time; all students in included groups are counted only once (Gross, 2012). The number of students

reported, "N" count, for the gap calculation is based on total school population and is not broken down grade by grade. Further, schools receive reporting information on individual gap groups, but accountability is based on the school's efforts at closing the combined non-duplicated gap group.

Growth Component of Unbridled Learning

One important component of the new accountability system is the Growth calculation, which is designed to measure a student's growth in learning from one year to the next, as compared to the student's academic peers. This is a measure educators in Kentucky have discussed for many years, eliminating the "apples to oranges" comparisons of the past (Tungate, 2010). This component measures how every individual student is making progress using student growth percentiles. The growth component recognizes schools and districts for the percentage of students demonstrating typical or higher levels of growth in reading and mathematics. Elementary and middle schools use annual reading and mathematics tests in grades three to eight to determine growth; high schools use PLAN at the 10^{th} grade level and ACT at the 11^{th} grade level to determine composite scores in reading and math as well as measure growth. The schools and districts are awarded points for the percentage of students showing typical or higher growth, which is defined as being at the 40^{th} percentile or above in relation to all students that started at the same percentile based on scores from the previous year.

College and Career Readiness Component of Unbridled Learning

The College and Career Readiness category measures how well schools and districts are preparing students for life after high school. This measure is be determined by the EXPLORE test for the middle schools and makes up 16% of the middle school accountability total score. At the high school level, the number of high school graduates who have successfully met one of the indicators for readiness for college and/or career is used to determine the college and career readiness measure for accountability, which makes up 20% of the high school accountability total score. The college ready indicator includes graduates who have met the Kentucky Council on Postsecondary Education (CPE) benchmarks for reading, English, and mathematics on any administration of the ACT. It also includes students who passed a college placement test like Compass or KYOTE. The career ready indicator includes graduates who have met benchmarks for Career-Ready Academic (ASVAB or ACT WorkKeys) and Career-Ready Technical (KOSSA or received an Industry-Recognized Career Certificate.) The college readiness percentage is determined by dividing the number of high school graduates who have met one of the indicators discussed above by the number of total graduates. In addition, high schools receive a half-point bonus for each graduate who meets the college AND career ready criteria.

Graduation Rate Component of Unbridled Learning

A structure known as the Average Freshman Graduation Rate (AFGR) is used to measure the Graduation Rate component of the accountability model. High schools and districts are evaluated on how many students are graduating on time. The United States

Department of Education has set forth expectations that all states utilize a cohort structure. However, until the cohort structure is available to Kentucky, the AFGR will be used to calculate the graduation rate for accountability at the high school level only. The Graduation Rate component is worth 20% of the accountability score at the high school level. See Table 2.1 for a summary of the measures used in each category and at each school level in accountability ratings.

Table 2.1

Accountability Categories within Next-Generation Learners

Grade Range	Achievement	Gap	Growth	College/Career Readiness	Graduation Rate
Elementary	Tests: Reading, mathematics, science, social studies, and writing	Tests: Reading, mathematics, science, social studies, and writing	Reading and Mathematics	N/A	N/A
Middle	Tests: Reading, mathematics, science, social studies, and writing	Tests: Reading, mathematics, science, social studies, and writing	Reading and Mathematics	EXPLORE (College Readiness)	N/A
High	End of Course Tests	End of Course Tests	PLAN to ACT Reading and Mathematics	College/Career Readiness Rate	AFGR/Cohort Model

Source: Kentucky Department of Education. (2012B, June). Next generation learners proposed accountability model: Draft for discussion. Office of Accountability.

Bonus Calculation in Unbridled Learning

When addressing the Bonus Calculation component of the Unbridled Learning Accountability Model, the Kentucky Board of Education requested that the accountability model not be designed in a manner that allows distinguished students and their scores to

mask the number of novice students. Therefore, each percent of students earning the distinguished label will also receive one-half point bonus, whereas the percent novice earns a negative half-point. When these bonuses are added, the novice may offset the distinguished. However, if the novice performance outweighs the distinguished, no points are added or subtracted.

Percentile Ranking in Unbridled Learning

The percentile ranking for schools and districts are based upon points from three to five of the categories described above. Achievement, gap, growth, college/career readiness, and graduation rate are weighted and then added for an overall score at the high school level. Graduation rates are not used at the elementary or middle school levels, and College and Career Readiness is not a factor at the elementary level. The weights of each category relevant to each specific school level are summarized in Table 2.2. Once these scores are compiled for schools across the state, they are rank-ordered and placed into percentiles, which determines a school/district overall category of distinguished, proficient, or needs improvement. Those schools or districts scoring in the top 90^{th} percentile from a particular level (elementary, middle, or high) receive the label of Distinguished. Those schools or districts scoring in the top 30^{th} percentile of all schools or districts in a particular level are labeled as Proficient. Finally, those schools or districts scoring at the 69^{th} percentile or below are labeled Needs Improvement.

Table 2.2

Weights for Unbridled Learning Components

Grade Range	Achievement	Gap	Growth	College/Career Readiness	Graduation Rate	Total
Elementary	30%	30%	40%	N/A	N/A	100%
Middle	28%	28%	28%	16%	N/A	100%
High	20%	20%	20%	20%	20%	100%

Source: Kentucky Department of Education. (2012B, June). Next generation learners proposed accountability model: Draft for discussion. Office of Accountability.

Rewards and Assistance Categories in Unbridled Learning

The Unbridled Learning Accountability model includes five reward/assistance categories to label schools meeting various criteria. The first reward category is a School or District of Distinction. These are high performing elementary, middle, or high schools or districts that meet current AMO; student participation rates on tests; the graduation rate goal, which is a rate above 60% for the prior two years; scores at the 95th percentile or higher on the overall accountability score; and for a district, does not have a school in the categories of focus or priority schools.

The second reward category in the Unbridled Learning Accountability Model is Highest Performing School or District. These are elementary, middle, or high schools or districts that meet current AMO; student participation rate on tests and the graduation rate goal. These schools also must have a graduation rate above 60% for the prior two years and score at the 90th percentile or higher on the overall accountability score. A district can reach this status when it does not have any schools categorized as Focus Schools or Priority Schools.

Another reward category for schools is High Progress Schools/Districts. This category includes both Title I and non-Title I schools showing the highest progress, as

compared to their peers. High Progress Districts also include those with the highest progress in comparison to their peers.

The assistance categories of the Unbridled Learning Accountability system begin with Focus School. To be categorized a Focus School, a school must have a non-duplicated student gap score in the bottom 10% of non-duplicated gap group scores for all elementary, middle, and high schools. Schools in this category also may have an individual student subgroup within assessment grades by level with a score in the third standard deviation below the average score for all students in the state. A school that has a graduation rate that has been below 60% for two consecutive years would also be categorized as a Focus School. A district can be categorized as a Focus District when it has a non-duplicated student group score in the bottom ten percent of non-duplicated student gap group scores for all districts.

The final category of assistance for Unbridled Learning is Priority School. A school is placed in this category when it has been labeled as persistently low achieving (PLA), which is defined by Kentucky Revised Statute KRS.160.346. This category will not apply to a district until the Unbridled Learning Accountability System has been in place for three consecutive years (Gross, 2012). Priority Schools "must document meaningful family and community involvement in the strategies for improvement outlined in their Continuous School Improvement Plans" (CSIP) and receive supports from the Kentucky Department of Education (Gross, 2009, p. 33). At this time, the supports for Priority Schools include Educational Recovery Teams, which generally consists of an Educational Recovery Leader, an Educational Recovery Literacy Specialist, and an Educational Recovery Mathematics Specialist. Schools have also

received additional funding to aide in the school improvement. Summative data on the numbers of schools and districts being classified in each reward and assistance category are presented in Table 2.3.

Table 2.3

Number of Schools and Districts by Rewards and Assistance Category Based on 2011-2012 Test Data

School Level	School/District of Distinction	Highest Performing School/District	Focus School/District	Priority School/District
Elementary	40	37	103	0
Middle	17	17	106	9
High	11	8	75	32
Totals	68	62	284	41
Districts	9	9	17	N/A

Source: Gross, L. (2012). First results from unbridled learning accountability model released. Kentucky Department of Education.

First Accountability Results from Unbridled Learning

On November 2, 2012, the Kentucky Department of Education released the first results from the Unbridled Learning Accountability System. These results indicated that, in 2011-2012, 47% of high school students were college and career ready. This represents a nine point gain from 2010-2011, when the college and career readiness rate was 38%, and a 13% increase from the 2009-2010 college and career readiness rate. College and career readiness was the highlight of the data release, with two-thirds of the schools and districts in the Needs Improvement category (Gross, 2012; Ujifusa, 2012). Dr. Terry Holliday, Kentucky Commissioner of Education, encouraged stakeholders to think of the results as a starting point. Holliday ensured stakeholders that the results were not indicating failure of our schools (Gross, 2012).

Explanation of Most Popular Accountability Models

The literature around accountability models consistently identifies four types of models that are predominantly used across the nation: status, improvement, growth, and value-added models (Goldschmidt, Roschewski, Choi, Auty, Hebbler, Blank, & Williams, 2005; Yu, Kennedy & Teddlie, 2007). Prior to *NCLB*, many states utilized status models. However, upon the enactment of *NCLB*, all state-approved accountability systems used status based approaches to evaluate their educational systems (Betebenner & Linn, 2009).

Status Accountability Models

The status accountability model is characterized as being a picture of an entity at one point in time. Observed proficiency levels are compared to an established target. AYP is an example of a status model (Yu et al., 2007; Zvoch &Stevens, 2008). Further, status models are "often contrasted to growth models…progress is defined by the percentage of students achieving at the proficient level for that particular year, and the school is evaluated based on whether the student group met or did not meet the goal." The basic question at work in a status model is "On average how are student performing this year?" (Goldschmidt et al., 2005, p. 3).

Improvement Accountability Model

Much like the status model, an improvement accountability models compares the change in status at two different points in time (Yu et al., 2007). The improvement model is a form of status model "which measures the change between different groups of

students." The basic question for an improvement model is "On average, are students doing better this year as compared to students in the same grade last year?" (Goldschmidt et al., 2005, p. 3).

Growth Accountability Model

The growth accountability model shares similarities with the status and improvement accountability models. However, the growth model uses cohorts of students over at least two points in time to determine progress (Auty, 2008; Carey& Manwaring, 2011; O'Malley, Murphy, McClarty, Murphy, &McBride, 2011; Yu et al., 2007). Goldschmidt et al. (2005) explain that the basic question for a growth model is "How much, on average, did students' performance change?" (p. 4). Betebenner & Linn (2009) suggested that growth models "have found favor as the preferred method for analyzing student achievement data for accountability purposes" (p. 3) because as Goldschmidt et al. (2005) suggested, growth models provide "a more concise picture of what is happening to students as they progress through a school" as compared to simply using a status model (p. 7).

Value-Added Accountability Models

Value-added models, also a form of growth models, are statistically superior to other models because of the formulas used to account for variances in factors such as student achievement and growth, family background, current class size, or teacher experience (Betebenner & Linn, 2009; Yu et al, 2007). Goldschmidt et al. (2005) explained that the basic question of the value-added model is "On average, did the

students' change in performance meet the growth expectation? And/or "By how much did the average change in student performance miss or exceed the growth expectation?" (p. 5).

Wiliam (2010) explains that in the United States, value-added approaches are largely being used to determine the "effects of individual teachers on student achievement" (p. 112). Value-added models are very popular because of their ability to link student assessment outcomes to education quality, a requirement of NCLB. Amrein-Beardsley (2008) clarified that in using value-added models "teachers are not given inappropriate credit for having a stellar set of students or penalized for having a difficult-to-teach class. Teachers, schools, and districts are simply evaluated on the value they have added to student learning" (p. 65). The most popular example of a value-added accountability model is Tennessee's Education Value-Added Assessment System (EVAAS), which is also being used in Ohio, Pennsylvania, and school districts across the nation (Betebenner & Linn, 2009).

Accountability Model Concerns

Linn (2006) contends that although value-added models provide "substantial improvements over the current-state and improvement for successive cohorts approaches to accountability," they still have limitations. For example, there is still a lack of definitive evidence "that school differences in student gains in achievement are attributable solely to differences in school quality," and they should not be used as "direct evidence" of instructional practice (p.19). Martineau (2006) suggests that "value-added models introduce remarkable distortions in the value-added estimates of the majority of

educators" by incorrectly identifying ineffective or effective teachers and schools and incorrectly attributing prior teacher and schools effects to later teachers and schools (p.35).

Goldschmidt et al. (2005) state that measuring growth like that measured by a value-added model is more expensive than other models because of its complexity and often more difficult to explain to stakeholders. Cost factors for growth models include infrastructure, setting the growth standards, availability of psychometric expertise, data system requirements, and training to build capacity among stakeholders. Further, states are required to adhere to AYP, and growth models are not allowed to mitigate the AYP rating (Goldschmidt et al., 2005). Ananda and Rabinowitz (2001) stated that status models like that of Texas might not be appropriate to use with schools that have a large variety of student performance because it may "unfairly penalize schools that demonstrate reasonable progress, but do not yet meet the common performance standards" (p. 7). "The recent changes in educational accountability have resulted in some state systems that have become overloaded in trying to serve too many purposes simultaneously, failing to serve any of them well" (Perie & Park, 2007, p. 5). For these reasons, a movement to "opt-out" of high stakes assessments has gained steam across in many states, including Washington, Maine, Colorado, New York, and Illinois (Marcus, 2012).

CHAPTER THREE

METHODS

Introduction

This study addressed a perceived need in the literature for research on the implications of different revisions of the same accountability system on the same student data. There was a specific need to determine if different versions of the same accountability systems would provide like results or provide results that differ and to what degree. One concern for schools and districts across Kentucky is the identification of the highest and lowest performing schools in the state. Because of the rewards, recognition, and consequences associated with such labels, there is a need to be very intentional in determining what a state accountability model will be designed to measure and choosing the model that most reliably supplies those data. Creating an accountability system that aligns with the design of the accountability model and provides data that best identify a school or districts' performance on the specific variable defined and consistently identifies the same high and low performers is crucial to the reliability of high stakes accountability models and systems.

Problem and Purposed Overview

The purpose of this study was to explore the implications of five different drafts of Kentucky's Unbridled Learning Accountability Model using the same student data. Another purpose of this study was to inform school, district, state, and national educators; policymakers; and other stakeholders to help ensure they clearly determine what

variables will be measured by their accountability systems and create the most effective assessments and accountability models possible to measure these determined variables. Further, this study provides research for the stakeholders to make them more aware of the implications of their work on districts and schools of differing demographics. This study also provides a rationale for being very intentional in the steps in creating an accountability system: first, be very intentional in determining what exactly will be measured; second, be intentional about which accountability model is most valid and reliable in measuring these variables; and third, be intentional and careful designing an accountability system that accurately and consistently measures school and district performance and consistently identifies the high and low performing schools and districts on the determined variables. Lastly, those designing accountability models also must determine how the results will be transparent to all stakeholders.

Rationale for Study

There are several studies addressing the validity, reliability, and other concerns with each of the different accountability models most widely used by states to determine student achievement, with each study offering suggestions for choosing a model (Betebenner & Linn, 2009; Choi, Goldschmidt, & Yamashiro, 2006; Yu et al., 2007), but these studies do not address how drafts of the same accountability model impact schools and districts. Another study similar to this study addressed which schools would be identified as lowest performing under a proposed revision of the *Elementary and Secondary Education Act* (McEachlin & Polikoff, 2012). That study utilized a more global approach to looking at the implications of national policies that impact the

development of state accountability systems, whereas this study focuses on the implications of one accountability system and the drafts leading to the final model upon the schools and districts of Kentucky.

Significance of Study

This study is significant because of the importance placed upon the success of current educational reform in America. The United States educational system is no longer ranked among the top internationally. Kress et al. (2011) suggest "a growing and widespread concern about the vital importance of education to our national security" (p. 188). If we are to succeed in reclaiming our mark among the top education systems in the world, we must accurately assess the success of our system. The success of our nation's educational system is primarily measured by high-stakes accountability tests, models, and systems. The study of the implications of different high-stakes accountability models utilizing the same student data is imperative to current education reform because we must ensure our educational systems are identifying the schools and districts that are best producing students with the 21st century skills needed to compete globally and ensure the economic success of the United States. Further, we must be able to accurately identify those schools and districts not producing students with these skills because research strongly suggests these students will not only most likely be required to take remedial courses if they pursue post-secondary education, but they will earn significantly less income over the course of their lifetime. The Hamilton Project reports that nearly 80% of high school dropouts made less than $30,000 in 2010, while 80% of college graduates earned around $100,000 (Greenhouse et al., 2012). Further, The McKinsey Global

Institute suggests there will be a global shortage of thirty-eight to forty million college educated workers by 2020. In addition, the group suggests there will be a potential shortage of forty-five million workers with the secondary education needed to be qualified for labor-intensive manufacturing and services (Dobb et al., 2012).

Limitations of Study

Limitations to this study are varied in nature. A few limitations revolve around the data available from Kentucky's new accountability model, Unbridled Learning. For example, the 2011-2012 school year was the first administration of the KPREP testing system used for the accountability model. Because this was the first administration, stakeholders were unsure of the alignment between the standards and the developed test. Another limitation for the state education department and school districts across the state includes the limited resources at this time of economic stress. Because of *No Child Left Behind* and Senate Bill 1, the Kentucky Department of Education developed an entirely new testing system and accountability model. School districts had very limited resources and time to take the steps necessary to quickly align curricula to a new set of standards and prepare for the new tests they knew very little about during the first administration. The transparency of the new accountability system was limited because psychometric formulas provided by KDE were limited and non-existent in some cases, leaving schools unable to fully understand their accountability results. Lastly, the results of this study were limited to the study of the accountability model for the state of Kentucky only.

Research Questions

1. To what extent would changes to Kentucky's Unbridled Learning Accountability Model result in different conclusions when utilizing the same student data?
2. Which school types are positively/negatively affected by each of the Kentucky models that were under consideration circa 2010?

Research question number one addresses changes in a schools reward of assistance categories schools could be placed in based on their accountability scores. These categories include the top and bottom five percent of schools, Schools and Districts of Distinction, Focus School, and Priority Schools. For the purpose of this research, the top and bottom five percent of elementary and high schools were analyzed using the total accountability scores. Research question number two sought to determine if schools with specific characteristics were positively or negatively impacted. These characteristics include school size, percentage of non-white students, percentage of students eligible for free and reduced lunch rates, per-pupil expenditure, average teaching experience, and teacher education.

Research Hypothesis

Kentucky's Accountability Model will result in different conclusions when utilizing the same student data and different drafts of the same model. Schools with diverse populations will be negatively affected by one of the proposed models more so than others. Furthermore, the models will identify different schools in the various rewards and assistance categories as define by the Unbridled Learning Accountability system.

Population Data

The data used for analyses included all elementary and high schools tested using the Kentucky Performance Rating of Educational Progress (K-PREP) during the 2011-2012 testing window. The data of 723 elementary and 230 high schools in Kentucky were used for the research analyses. According to the State Report Card, available online, 649,688 students were enrolled in Kentucky's public schools during the 2011-2012 school year. The elementary schools served 152,121 student scores, while the high schools in the data set enrolled 47,880 student scores. Kentucky's public schools are comprised of approximately 82% Caucasian, 10% African-American, 1.3% Asian, and 3.8% Hispanic students (Kentucky Department of Education, 2011A; Kentucky Department of Education, 2012).

Description of Research Design

The dependent variable in this study is the overall school index score resulting from the 2011-2012 elementary and high school K-PREP test data, which is available to the public on the Kentucky Department of Education website. The independent variables for this study include various considered drafts of the K-PREP accountability model as outlined in this section. Scores for each component of the two high schools models and three elementary models were calculated. Bivariate correlations were calculated to assess the relationships between the overall scores in each model, as well as the relationship between the various components within the adopted model. The influences of school, student, and teacher characteristics on model results are also reported. The top and bottom five percent of schools for each of the models in the study were calculated

because the rewards and assistance categories for the Unbridled Learning Accountability System revolve around the bottom and top five percent of schools.

Overall K-PREP Information

The collective tests used for assessment and accountability purposes are referred to as K-PREP. NCS Pearson has developed the norm and criterion referenced tests for grades three through eight and the on-demand writing portion of the high school testing. ACT *Quality Core* provides the end-of-course (EOC) assessments at the high school for English II, Algebra II, U.S. History, and Biology. Each of these tests included multiple choice and constructed response sections during the 2011-2012 administration of the KPREP. The EOC assessments are given at the conclusion of each of the courses. The remaining K-PREP tests are given over a window of five days and within two weeks of the close of the school year.

Elementary K-PREP Information

As Table 3.1 illustrates, third graders across the state of Kentucky test in reading and mathematics. Reading and mathematics assessments include three parts, with the first part being the norm-referenced test (NRT). The remaining parts of the test are criterion-referenced tests (CRT). The standards being tested in reading and mathematics are the Kentucky Core Academic Standards (KCAS). Students are required to read passages, answer multiple choice questions, and complete short answer questions. Third graders are tested for the K-PREP test for a total of 235 minutes.

Table 3.1

3rd Grade K-PREP Testing Information

Reading	#Psg	#MC	#SA	Time	**Math**	#MC	#SA	Time	
Part A-NRT		30		40	Part A-NRT	30		40	
Part B	2	12	1	35	Part B	30	3	45	
Part C	3	18	2	50	Part C	11	2	25	
Reading total time					125 Min	Math total time			235 Min
MC=Multiple Choice, SA= Short Answer, ER= Extended Response, Psg= Passage									

Source: Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times.

As Tables 3.2 and 3.3 demonstrate, fourth graders across the state of Kentucky test in reading, mathematics, science, and language mechanics, and are tested a total of 460 minutes. The first section of each of the content tests is the norm-referenced test. The remaining sections of the test are criterion-referenced tests. The standards being tested in reading, math, and language mechanics are also KCAS, while science are Kentucky's Core Content Science Standards. Students are required to read passages, answer multiple choice questions, complete short answer questions, and complete extended response questions.

Table 3.2

4th Grade K-PREP Reading and Math Testing Information

Reading	#Psg	#MC	#SA	#ER	Time	**Math**	#MC	#SA	#ER	Time
Part A NRT		30			40	Part A-NRT	30			40
Part B	2	12	1	1	45	Part B	26	3	2	75
Part C	3	18	2	1	60	Part C	12	2		25
Total Testing Time Reading					145	Total Testing Time Math				140
MC=Multiple Choice, SA= Short Answer, ER= Extended Response, Psg= Passage										

Source: Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times.

Table 3.3

4th Grade K-PREP Science and Language Mechanics Testing Information

Science	#MC	#ER	Time	**Language Mechanics**	#MC	Time
Part A- NRT	30		40	Part A- NRT	30	40
Part B	21	2	55			
Part C	21	1	40			
Total Testing Time Science			135	Total Testing Time Language Mechanics		40
MC=Multiple Choice, SA= Short Answer, ER= Extended Response, Psg= Passage						

Source: Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times.

Fifth graders are tested in reading, mathematics, social studies, and on-demand writing. As displayed by Tables 3.4 and 3.5, the first section of each of the reading, mathematics, and social studies tests is the norm-referenced test. The remaining sections of the tests are criterion-referenced tests. The standards being tested in reading, math, and on-demand writing are KCAS. The standards being addressed in social studies are the Kentucky's Core Content. Students are required to read passages, answer multiple choice questions, complete short answer questions, complete extended responses, and complete on-demand writings (Office of Assessment and Accountability, 2012).

Table 3.4

5th Grade K-PREP Reading and Math Testing Information

Reading	#Psg	#MC	#SA	#ER	Time	**Math**	#MC	#SA	#ER	Time
Part A-NRT		30			40	Part A-NRT	30			40
Part B	2	17	1	1	50	Part B	30	3	2	75
Part C	3	24	2	1	65	Part C	13	2		25
Total Testing Time Reading					155	Total Testing Time Math				140
MC=Multiple Choice, SA= Short Answer, ER= Extended Response, Psg= Passage										

Source: Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times.

Table 3.5

5th Grade K-PREP Social Studies and On-Demand Testing Information

Social Studies	#MC	#ER	Time	**On-Demand Writing**	# Stand Alone	# Psg Based	Time
Part A- NRT	30		40	Part A	1		30
Part B	22	1	55	Part B		1	90
Part C	28	1	40	Total Time On-Demand Writing			120
Total Testing Time Social Studies			135				

MC=Multiple Choice, SA= Short Answer, ER= Extended Response, Psg= Passage

Source: Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times.

High School K-PREP Information

At the high school level, 10th and 11th graders are tested on on-demand writing, which is a criterion-referenced test designed to assess the KCAS writing and language standards. Table 3.6 shows the time allotted and describes the type of on-demand given at grades 10 and 11. Tenth graders also take the PLAN test, a test created by ACT that measures a students' progress toward ACT benchmarks.

Table 3.6

High School K-PREP On-Demand Testing Information

Grade Level	On-Demand Writing	# Stand Alone	#Psg Based	Time
10th Grade	Part A	1		40
	Part B		1	90
11th Grade	Part A	1		40
			1	90

Source: Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times.

Table 3.7 is a blueprint for the PLAN test. Eleventh graders take the ACT, which measures a students' academic career readiness for college. Table 3.8 highlights

the makeup of the ACT and the time allotted for each content area of the test. ACTS' Quality Core end-of-course assessments are also given at the end of English II, Algebra II, U.S. History, and Biology.

Table 3.7

10th Grade K-PREP PLAN Testing Information

Subject	**Number of Questions**	**Time Allowed**
English		30 minutes
Usage/Mechanics	30	
Rhetorical Skills	20	
Math		40 minutes
Pre-Algebra/Algebra	22	
Geometry	18	
Reading	25	20 minutes
Science	30	25 minutes

Source: Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times.

Table 3.8

11th Grade K-PREP ACT Testing Information

Test	**#MC Questions**	**Time**
English	75	45 minutes
Mathematics	60	60 minutes
Reading	40	35 minutes
Science	40	35 minutes

Source: -ACT, Inc. (2013). Description of the ACT.

Table 3.9 reports the number of multiple choice and constructed response for each of the end-of-course exams. The end-of-course tests and on-demand writing are used to measure Achievement and Gap in the Next-Generation Learners accountability model. Kentucky uses the growth from PLAN to ACT to measure the Growth component of the accountability model.

Table 3.9

K-PREP End-of-Course Testing Information

Test	#MC Sessions	Number of Questions	Time Per MC Session	Constructed Response
English II	2	35-58	45 minutes	1
Algebra II	2	35-58	45 minutes	3
U.S. History	2	35-58	45 minutes	2
Biology	2	35-58	45 minutes	3

Source: Kentucky Department of Education. (2013). Kentucky's high school end-of-course assessments: Answers to parents' most frequently asked questions.

Data Collection and Instrumentation

Existing databases made available to the public by Kentucky's Department of Education were used as the sources of data for this study. Excel spreadsheets containing overall KPREP scores and scores for each accountability component were downloaded from KDE's website. Data on school characteristics were downloaded from KDE's website as well. All files contained a unique school identification number that allowed all files to be merged into a single file. PASW version 21.0 statistical software was used to complete the data analysis to determine the variability between each of the drafts for Kentucky's elementary and high school accountability models.

Three proposed drafts for the elementary model were chosen and are highlighted in Table 3.10. Although model A is the adopted model, models B and C were also proposed within the development of the Unbridled Learning Accountability Model as seen in the Next-Generation Learners Proposed Accountability Model White Papers (See Appendix A; Draut, 2013; Kentucky Department of Education, 2011C; Kentucky Department of Education, 2011D, Kentucky Department of Education, 2011E, Kentucky Department of Education, 2012B; 2011d; Sims, 2012).

Table 3.10

Weighting of Components within Models Proposed for K-PREP Elementary Level

Elementary Models	Achievement	Gap	Growth
Model A	30%	30%	40%
Model B	25%	25%	50%
Model C	33.33 %	33.33.%	33.33%

Note: Model A is the adopted model for Kentucky's Unbridled Learning Accountability Model.

At the high school level, only two drafts were chosen for the study. Model A is the adopted model for the high school accountability for Unbridled Learning. Although Model B is not an official draft model, it does reflect the deep commitment and focus Kentucky has placed on college and career readiness as seen in the Next-Generation Learners Proposed Accountability Model White Papers (See Appendix A).Table 3.11 presents each of the five components at the high school level and their corresponding percentages for accountability.

Table 3.11

Weighting of Components within Models Proposed for K-PREP High School Level

High School Model	Achievement	Gap	Growth	College/Career Readiness	Graduation Rate
Model A	20%	20%	20%	20%	20%
Model B	15%	15%	15%	40%	15%

Note: Model A is the adopted model for Kentucky's Unbridled Learning Accountability Model.

CHAPTER FOUR

RESULTS

The primary purpose of this chapter is to report the findings of this research study. This chapter revolves around the two research questions. Specifically, the following results emerged from the analyses to address the following research questions:

1. To what extent would changes to Kentucky's Accountability Model result in different conclusions when utilizing the same student data?
2. Which school types are positively/negatively affected by each of the models that were under consideration?

Research question number one addresses changes in the reward or assistance categories schools could be placed in based on their accountability scores. Specifically, the categories of the top and bottom five percent of schools are analyzed. Research question number two sought to determine if schools with specific characteristics were positively or negatively impacted. These characteristics include school size, percentage of non-white students, percentage of students eligible for free and reduced lunch rates, per-pupil expenditure, average years of teaching experience, and teacher education.

Participants

The data were drawn from those of all elementary and high schools tested using the Kentucky Performance Rating of Educational Progress (K-PREP) during the 2011-2012 testing window. The data for 723 elementary and 230 high schools in Kentucky

were used to complete the what-if kind of analysis reported in the next chapter. According to the State Report Card available online, 649,688 students were enrolled in Kentucky's public schools during the 2011-2012 school year. The study included the accountability data from a total of 174 school districts with a total of 1,233 schools. The elementary data set includes schools that enrolled 152,121 students, while the high school data set includes 47,880 students. Kentucky's public schools are comprised of approximately 82% Caucasian, 10% African-American, 1.3% Asian, and 3.8% Hispanic students (Kentucky Department of Education, 2011A).

Research Findings

Three Elementary Models

Question one focused on the extent to which Kentucky's Accountability Model at the elementary level would result in different conclusions when utilizing the same data. As an initial analysis, bivariate correlations were run to assess the relationships between the three total scores that emerged from each model. As noted in Table 4.1, the relationships among overall proficiency scores of all three models were statistically significant with exceptionally high positive correlations. In fact, the lowest correlation was between model B and model C $r(723)=.982$, which is still indicative of a near perfect linear relationship.

Table 4.1

Correlations between Elementary Models A, B, and C

Model Name	Model A	Model B	Model C
Model A- Pearson Correlation	1	.994	.997
Sig. (2-tailed)		.000	.000
N	723	723	723
Model B- Pearson Correlation	.994	1	.982
Sig. (2-tailed)	.000		.000
N	723	723	723
Model C- Pearson Correlation	.997	.982	1
Sig. (2-tailed)	.000	.000	
N	723	723	723

Given the extremely high correlations between the total scores in each model, it was critical to analyze the correlations between the three components within each model. These correlations are represented in Table 4.2. Lower correlations between model components indicate differences between component scores could be masking differences in total scores between models. The highest statistically positive correlation is between achievement score and the total score for model A, $r(723)=.920$. The significantly high correlation between model A's achievement and gap scores $r(723)=.880$ are so highly correlated that one could conclude the achievement score is in effect the same score as the gap score. The correlations of the growth score with the achievement score $r(723)=.556$ and the gap score $r(723)=.425$ were moderate but significant as well.

Table 4.2

Correlations between Elementary Model A Achievement, Gap, and Total Score

		Achievement Score	Gap Score	Growth Score	Model A Total Score
Achievement Score	Pearson Corr.	1	.880	.559	.920
	Sig. (2-tailed)		.000	.000	.000
	N	723	723	723	723
Gap Score	Pearson Corr.	.880	1	.425	.851
	Sig. (2-tailed)	.000		.000	.000
	N	723	723	723	723
Growth Score	Pearson Corr.	.559	.425	1	.815
	Sig. (2-tailed)	.000	.000		.000
	N	723	723	723	723
Model A Total Score	Pearson Corr.	.920	.851	.815	1
	Sig. (2-tailed)	.000	.000	.000	
	N	723	723	723	723

After the correlations were calculated, total accountability scores were calculated based on elementary statewide achievement data for models A, B, and C. The cut scores for the top and bottom five percent for each of the models were identified and are reported in Table 4.3. As a reminder, Model A was the adopted model. All results in Models B and C are within one point of the scores in Model A.

Given the minimal differences between cut scores, it is not surprising that changes in the elementary model A, B, and C rankings and total accountability scores were minimal for both the top and bottom 5% of elementary schools. See Tables 4.4 and 4.5 for these rankings and scores.

Table 4.3

Cut Scores for Top and Bottom Five Percent of Schools under Elementary Models A, B, and C

Percentiles	Model A	Model B	Model C
Bottom 5%	42.30	42.73	41.56
Top 5%	72.58	73.02	72.59

Table 4.4 displays the total accountability score for the bottom five percent of schools for each of the three elementary models. Although the rankings and total accountability scores differed within each of the three models, 30 of the 37 schools identified in Model A were also identified within Models B and C. All of the 37 schools identified in Model A were identified in at least one of the other models. Six schools were identified in only one of the three models: Cordia, Fulton County, Salyersville, Big Creek, Majestic Knox Center, and Flat Lick. Because of rounding, only 36 schools were identified in the bottom 5% in Models B and C, where 37 schools were identified in Model A. Lastly, Models A and B are the most similar with only four schools different between the two models.

Table 4.4

Total Accountability Score for Elementary Schools in Bottom Five Percent Rankings under Models A, B, and C

Model A		Model B		Model C	
1. William H Natcher	42.3	Cordia	42.7	Flat Lick	41.5
2. Sanders	42.3	Sanders	42.6	Cumberland	41.4
3. Blaine	42.2	Fulton County	42.5	Blaine	41.1
4. Shelby Trad.	42.0	Eminence	42.1	Shelby Trad.	41.2
5. Cumberland	41.9	Gutermuth	41.9	Estes	41.0
6. Oneida	41.7	Salyersville	41.5	Oneida	40.8
7. Mill Creek	41.2	Cardinal Valley	41.4	Cardinal Valley	40.7
8. Eminence	41.1	Lincoln	41.4	Green Hills	40.6

Table 4.4 (continued)

Model A		Model B		Model C	
9. Maupin	41.0	John G Carlisle	41.3	Eminence	40.5
10. Cardinal Valley	41.0	Big Creek	41.2	Mason-Corinth	40.4
11. Estes	40.3	Roosevelt Perry	41.1	Mill Creek	40.0
12. Gutermuth	40.2	R E Stevenson	41.0	R E Stevenson	39.7
13. R E Stevenson	40.2	Majestic Knox Creek	40.9	Maupin	39.4
14. Green Hills	39.8	Silver Grove	40.9	Fulton Ind.	39.2
15. Mason-Corinth	39.8	King	40.5	Gutermuth	39.0
16. John G Carlisle	39.5	Wheatley	40.4	Rousseau	39.0
17. Lincoln	39.4	Booker T Washington	40.1	Booker T Washington	38.7
18. Wheatley	39.3	Newport Inter.	39.9	Wheatley	38.6
19. Booker T Washington	39.2	William H Natcher	39.7	John G Carlisle	38.4
20. Roosevelt Perry	39.1	Cochran	39.5	Deming	38.0
21. Fulton Ind.	39.1	Estes	39.4	Lincoln	37.9
22. Deming	38.5	Deming	39.3	Roosevelt Perry	37.8
23. Silver Grove	38.2	Mason-Corinth	38.9	King	36.5
24. King	38.1	Fulton Ind.	38.8	Lewis Central	36.5
25. Rousseau	37.8	Green Hills	38.6	Silver Grove	36.4
26. Newport Inter.	37.8	Beckham Bates	38.0	Newport Inter.	36.3
27. Cochran	37.2	Semple	37.8	Owensboro Middle School South	36.1
28. Beckham Bates	36.7	Owensboro Middle School South	36.6	Semple	35.7
29. Semple	36.6	Lewis Central	36.5	Cochran	35.7
30. Lewis Central	36.5	Rousseau	35.6	Elkhorn City	34.0
31. Owensboro Middle South	36.3	Jacob	34.4	James A Cawood	32.6
32. Jacob	32.9	Goose Rock	32.02	Paces Creek	32.23
33. Elkhorn City	32.7	James A Cawood	31.25	Jacob	31.96
34. James A Cawood	32.1	Elkhorn City	30.9	Goose Rock	31.39
35. Goose Rock	31.7	Paces Creek	29.67	Chavies	28.09
36. Paces Creek	31.3	Chavies	27.75	Beckham Bates	35.79
37. Chavies	28.0				

Table 4.5 presents the mean accountability score for the top 5% of schools for each of the three elementary models. Like Table 4.4, the rankings and total accountability scores are similar with 29 of the 36 schools identified in Model A also identified in Models B and C. Two of the remaining 36 schools identified in Model A, the adopted model, were also identified in Model B or C. There were only eight schools identified in a single model: Johnson, Liberty, Northern, Highland, North Jackson, Highland, W R Castle Memorial, and Meade Memorial. Due to rounding, Model A identified 36 schools, Model B identified 31 schools, and Model C identified 33 schools for the top 5% of schools.

Table 4.5

Total Accountability Score for Elementary Schools in Top Five Percent Rankings under Models A, B, and C

Model A		Model B		Model C	
1. Greathouse Shryock	82.5	Greathouse Shyrock	82	Greathouse Shyrock	82.7917
2. Pilot View	81.5	Pilot View	80.95	Pilot View	81.8585
3. Jones Fork	81.2	Veterans Park	80.825	Jones Fork	81.7252
4. Veteran's Park	80.5	Jones Fork	80.475	Veterans Park	80.2586
5. Brandeis	79.8	Brandeis	80.175	Brandeis	79.625
6. May Valley	77.1	Lowe	77.625	May Valley	78.8921
7. Lowe	77.1	Benton	77.325	Benton	76.8923
8. Benton	77.0	Hager	76.7	Lowe	76.7923
9. Hager	76.3	Moyer	76.625	Scapa	76.1257
10. Goshen at Hillcrest	76.3	Goshen at Hillcrest	76.5	Goshen at Hillcrest	76.0257
11. Moyer	76.2	Rosa Parks	76.5	Hager	75.9591
12. Johnson	76.0	Mapleton	76	Moyer	75.8258
13. Rosa Parks	75.9	Trapp	75.725	Rosa Parks	75.6258
14. Scapa	75.8	Picadome	75.6	Trapp	75.0592
15. Trapp	75.3	River Ridge	75.575	North Pointe	74.4926

Table 4.5 continued

Model A		Model B		Model C	
16. North Pointe	74.9	Scapa	75.425	Beechwood	74.3592
17. River Ridge	74.6	North Pointe	75.4	Stopher	74.2592
18. Liberty	74.5	Stopher	74.85	Concord	74.259
19. Stopher	74.5	Providence	74.7	Providence	74.2259
20. Providence	74.4	Norton	74.525	River Ridge	73.9259
21. Picadome	74.4	May Valley	74.45	Picadome	73.5926
22. Northern	74.3	Shopville	74.275	Reidland	73.4593
23. Mapleton	74.3	Concord	74.25	Harmony	73.3593
24. Concord	74.3	Carter	74.025	Glendover	73.2593
25. Beechwood	73.9	Glendover	74.025	Mapleton	73.226
26. Shopville	73.6	Harmony	73.225	Carter	73.1927
27. Glendover	73.5	Star	73.225	Shopville	73.0594
28. Carter	73.5	Audobon Traditional	73.125	Auburn	72.7927
29. Norton	73.4	North Jackson	73.125	Shirley Mann	72.7927
30. Highland	73.3	Beechwood	73.075	W R Castle Memorial	72.7261
31. Harmony	73.3	Reidland	73.05	Norton	72.6594
32. Eastern	73.2			Meade Memorial	72.5927
33. Reidland	73.2			Star	72.5927
34. Star	72.9				
35. Shirley Mann	72.9				
36. Audubon Traditional	72.6				

The Two High School Models

To explore research question number one at the high school level, to what extent would changes to Kentucky's Accountability Model result in different conclusions when utilizing the same data, a bivariate correlation was run to assess the relationships between the two total accountability scores that emerged from the two models. As noted in Table 4.6, as was the case with the three elementary models, the relationship of Model A, the adopted model, with Model B is statistically significant and an exceptionally high positive correlation, $r(230)=.978$, $p<.001$.

Table 4.6

Correlation between High School Models A and B

Total Accountability Score	Model A	Model B
Model A Pearson Correlation	1	.978
Sig. (2-tailed)		.000
N	230	230
Model B Pearson Correlation	.978	1
Sig. (2-tailed)	.000	
N	230	230

Because of the extremely high correlations between the total scores for each of the high school models, it was critical to analyze the correlations between the three components within each model as seen in Table 4.7. Lower correlations between model components could indicate differences between component scores could be masking differences in total scores between models. The highest statistically positive correlation is between achievement score and the total score for model B, $r(230)=.926$. The significantly high correlation between model B's total score, achievement, gap, and

growth scores are so correlated that the achievement score is in effect the same score as the gap score and is highly correlated with the growth score as well.

Table 4.7

Correlations between High School Model A Achievement, Growth, Gap, and Total Score

		Achieve-ment Score	Gap Score	Growth Score	CCR Score Model A	Grad-uation Score Model A	Model A Total Score
Achieve-ment Score Model A	Pearson Corr. Sig.(2-tailed) N	1 230	.878 .000 230	.625 .000 230	.692 .000 230	.442 .000 230	.926 .000 230
Gap Score Model A	Pearson Corr. Sig. (2-tailed) N	.878 .000 230	1 230	.526 .000 230	.534 .000 230	.359 .000 230	.821 .000 230
Growth Score Model A	Pearson Corr. Sig. (2-tailed) N	.625 .000 230	.526 .000 230	1 230	.492 .000 230	.134 .043 230	.676 .000 230
College and Career Readiness Model A	Pearson Corr. Sig. (2-tailed) N	.692 .000 230	.534 .000 230	.492 .000 230	1 230	.412 .000 230	.857 .000 230
Graduation Score Model A	Pearson Corr. Sig. (2-tailed) N	.442 .000 230	.359 .000 230	.134 .000 230	.412 .000 230	1 230	.597 .000 230
Model A Total Score	Pearson Corr. Sig. (2-tailed) N	.920 .000 230	.821 .000 230	.676 .000 230	.857 .000 230	.597 .000 230	1 230

After high school model correlations were determined, total scores were calculated for all high schools under models A and B. Cut scores for the top and bottom five percent for each of the models were identified and are presented in Table 4.8. Cut scores differed between the two models by approximately 3.3 points.

Table 4.8

Cut Score for Top and Bottom Five Percent for High School Models A and B

Percentiles	High School Model A	High School Model B
Bottom 5%	41.47	38.1903
Top 5%	71.2415	67.94

Table 4.9 shows the total accountability scores for the bottom five percent of schools in each of the high school models. Although the same eleven schools are in the bottom five percent, rankings and scores did change as a result of the differences between high school models A and B. As a reminder, Model A was adopted by the Kentucky Board of Education as the accountability model. After reviewing Table 4.8 the total scores under model A are lower than those under model B. For example, the score for Perry Central for Model A is 37.4, while under model B, the score is 41.3, a difference of 3.9 points. The lower scores under the model highlight two key issues. First, if model B had been adopted, high schools would be closer to the goal score of 100. Second, given that model B doubles the College Career Readiness weighting and the total score increases compared to model A, this suggest that high schools in Kentucky are performing better in this area than the other four components of the model.

Table 4.9

Mean Total Accountability Score for Bottom Five Percent Rankings for High School Models A and B

School Name	Model A Score	School Name	Model B Score
1. Perry Central	41.3	1. Perry Central	37.4
2. Southern	41.2	2. Southern	37.3
3. Phelps	41.1	3. Western	34.8
4. Caverna	40.6	4. Caverna	34.7
5. Western	40.3	5. Phelps	34.0
6. Cordia	39.6	6. Cordia	33.8
7. Holmes	36.1	7. Holmes	32.3
8. Doss	35.8	8. Iroquois	32.3
9. Iroquois	34.4	9. Doss	30.3
10. Valley	31.0	10. Valley	26.0
11. The Academy @ Shawnee	27.9	11. The Academy@Shawnee	24.6

Table 4.10 presents the accountability score for the top five percent of schools for both high school models. Although nine schools are in the top five percent for both models, South Warren and South Oldham are not in the top five percent in model A and were replaced by Painstville and Hickman in Model B.

Table 4.10

Total Accountability Score for Top Five Percent Rankings for High School Models A and B

School Name	Model A Score	School Name	Model B Score
1. Dupont Manual	87.3	1. Dupont Manual	90.5
2. Beechwood	84.5	2. Beechwood	85.6
3. Highlands	79.0	3. Walton Verona	82.0
4. Walton-Verona	77.9	4. Highlands	78.5
5. Louisville Male	77.0	5. Louisville Male	77.2
6. North Oldham	76.0	6. North Oldham	77.0
7. Brown	75.7	7. Brown	75.9
8. Model Lab	73.5	8. Murray	73.3
9. Murray	69.6	9. Paintsville	73.1
10. South Oldham	68.9	10. Hickman	72.8

Table 4.10 (continued)			
11. South Warren	68.1	11. Model Lab	72.6

The Impact of Elementary Models on Rankings of Diverse Schools

The second research question sought to determine which school types are positively/negatively affected by each of the proposed elementary models. School characteristics analyzed included total enrollment, percentage of non-white students, percentage of low income students, average daily attendance, and per pupil spending. Teacher characteristics included average years of teaching experience and percentage of teachers with a Masters degree or above. Bivariate correlations were calculated to analyze the relationships between the school and teacher characteristics with total accountability scores under each proposed model. The results are presented in Table 4.11. Collectively, the results indicate that the correlations between the school and teacher characteristics with total accountability scores are very similar across each model.

The more important finding gleaned from Table 4.11 is that every school and teacher characteristic is significantly correlated with the total accountability scores. Specifically, there are statistically significant positive relationship between the total student enrollment, average years of teaching experience, and percentage of teachers with a Masters or above with the total accountability scores. As the total student enrollment, average years of teaching experience, and percent of teachers with a Masters degree or above increased, the total accountability scores also increased. On the contrary, under all three models, there are statistically significant negative relationships between percentage of non-white students, percentage of students eligible for free or reduced lunch, and spending per pupil with total accountability scores. In other words, there is a negative

impact on the accountability score based on these characteristics. As the number of non-white students, students eligible for percent free and reduced lunch, and per pupil spending increased, the accountability scores declined. As noted previously, these correlations were very similar across all three models, so it can be concluded that the three elementary models are unlikely to rank schools differently based on differences in these school and teacher characteristics. However, schools with more low income and non-white students, and schools that spend more per pupil are more likely to be rated lower under all three models.

Table 4.11

Elementary Accountability Score Correlation with School and Teacher Characteristics

School and Teacher Characteristic	Test Characteristics	Model A	Model B	Model C
Total Student Enrollment	Pearson Correlation	.178	.191	.167
	Sig. (2-tailed)	.000	.000	.000
	N	720	720	720
Percentage of Non-White Students	Pearson Correlation	-.253	-.221	-.273
	Sig. (2-tailed)	.000	.000	.000
	N	723	723	723
Percent Free or Reduced Lunch	Pearson Correlation	-.536	-.529	-.536
	Sig. (2-tailed)	.000	.000	.000
	N	723	723	723
Spending Per Pupil	Pearson Correlation	-.277	-.264	-.284
	Sig. (2-tailed)	.000	.000	.000
	N	721	721	721
Average Years Teaching Experience	Pearson Correlation	.197	.182	.206
	Sig. (2-tailed)	.000	.000	.000
	N	722	722	722
Percent of Teachers with a Masters or Above	Pearson Correlation	.148	.137	.154
	Sig. (2-tailed)	.000	.000	.000
	N	723	723	723

Table 4.12 presents data highlighting the relationships between total elementary accountability scores and school characteristics, with the accountability scores broken

down into deciles. Comparing the highest and lowest performing deciles reveals stark contrasts. For example, the mean percentage of students eligible for free/reduced lunch for the bottom decile (M=81.4) is double the percentage in the top performing decile (M=41.1). Similarly, the schools in the top decile enroll a percentage of non-white students (M=15.1) that is less than half of the same percentage in schools in the bottom decile (M=32.4). Finally, while it may be initially surprising that schools ranked in lower deciles spend more per pupil on average than schools in the upper deciles, this finding is likely the result of additional funds such as Title I that are allocated for low income students.

Table 4.12

School Characteristics by Decile Accountability Scores under Model A

Total Accountability Score Under Model A	Mean Total School Enrollment	Mean Percent Free or Reduced Lunch	Mean Percentage of Non-White Students	Mean Average Daily Attendance	Mean Spending Per Pupil
Bottom Decile	378.46	81.41	32.38	367.36	$9,934.06
2nd Decile	388.09	78.08	25.81	369.26	$8870.50
3rd Decile	413.64	70.30	17.50	393.11	$8,410.72
4th Decile	434.99	70.43	18.14	415.02	$8,301.89
5th Decile	423.79	67.74	16.35	404.03	$8,041.27
6th Decile	428.82	65.43	15.27	410.91	$8,037.48
7th Decile	479.24	52.92	13.23	460.89	$7,588.87
8th Decile	465.04	53.91	13.65	447.11	$7,800.73
9th Decile	426.38	55.61	10.63	410.12	$8,003.31
Top Decile	482.38	41.07	15.14	463.75	$7,977.72
Total	432.16	63.66	17.79	414.17	$8,294.28

Table 4.13 highlights the relationships between total elementary accountability scores and teacher characteristics, with the accountability scores broken down into deciles. As average years of teaching experience and percentage of teachers with a

Masters degree or above increases, the accountability score decile also increases. These findings support investments in furthering teacher education levels and the importance of retaining experienced teachers. They also support the importance of placing such teachers in the schools that serve students with the greatest needs.

Table 4.13

Mean Accountability for Elementary Model A with Teacher Characteristics by Decile

Accountability Scores Model A	Average Years Teaching Experience	Percent of Teachers with a Masters or Above
Bottom Decile	10.566	74.615
2nd Decile	11.606	79.366
3rd Decile	11.497	80.250
4th Decile	11.407	78.944
5th Decile	11.758	78.239
6th Decile	11.781	80.169
7th Decile	12.431	83.073
8th Decile	11.992	82.081
9th Decile	12.060	80.807
Top Decile	12.789	81.965
Total	11.786	79.955

Characteristics of High Schools Receiving Rewards or Sanctions under Different Models

The second research question for this study sought to determine the characteristics of high schools that are positively/negatively affected by each of the models presented. Bivariate correlations were conducted to determine the relationships of the total accountability scores for the high schools with percent of non-white students, total enrollment, percentage of students eligible for free and reduced lunch, spending per pupil, average years teaching experience, and percentage of teachers with a Masters degree or above. These correlations were calculated separately for both high school models. The results are presented in Table 4.14.

Table 4.14

High School Accountability Score Correlation with School and Teacher Characteristics

School and Teacher Characteristic	Test Characteristics	Model A	Model B
Total Student Enrollment	Pearson Correlation	.174	.175
	Sig. (2-tailed)	.000	.000
	N	230	230
Percentage of Non-White Students	Pearson Correlation	-.181	-.194
	Sig. (2-tailed)	.000	.000
	N	230	230
Percent Free or Reduced Lunch	Pearson Correlation	-.647	-.638
	Sig. (2-tailed)	.000	.000
	N	230	230
Spending Per Pupil	Pearson Correlation	-.213	-.212
	Sig. (2-tailed)	.000	.000
	N	230	230
Average Years Teaching Experience	Pearson Correlation	.264	.251
	Sig. (2-tailed)	.000	.000
	N	230	230
Percent of Teachers with a Masters or Above	Pearson Correlation	.235	.208
	Sig. (2-tailed)	.000	.000
	N	230	230

In models A and B, all correlations were statistically significant. As was the case with the elementary schools, positive relationships were found between the total student enrollment, average years of teaching experience, and percent of teachers with a Masters or above with the total accountability score. As the total student enrollment, average years of teaching experience and percent of teachers with a Masters degree or above increased, the total accountability score also increased. However, in models A and B, there were statistically significant negative relationships between percentage of non-white students, percentage of students qualifying for free or reduced lunch, and spending per pupil with total accountability scores. As the percentage of non-white students, percentage of students eligible for free and reduced lunch, and spending per pupil

increases, the accountability score decreases. Both of the high school models yielded similar positive and negative relationships between accountability scores and school characteristics. Therefore, it can be concluded the high school models do not appear to be influencing high schools with different school characteristics in different ways. Under both models, schools with high percentages of low income students are the ones the most likely to receive lower accountability scores. Race and per pupil spending expenditures also are correlated negatively with total accountability scores.

Table 4.15 highlights the relationships between total high school accountability scores and school characteristics by dividing accountability scores into deciles. The deciles reveal similar patterns to those found in the elementary school deciles. Specifically, schools earning accountability scores in the bottom decile enroll over twice the percentage of low income (M=68.7) and non-white (M=26.8) students compared to schools in the top percentile, which serve fewer low income (M=29.4) and non-white students (M=12.6). High schools in the lower deciles also spend significantly more dollars per pupil.

Table 4.15

Total High School Accountability Scores in Deciles with Mean School Characteristics

Total Accountability Score Under Model A	Mean Total School Enrollment	Mean Percent Free or Reduced Lunch	Mean Percentage of Non-White Students	Mean Average Daily Attendance	Mean Spending Per Pupil
Bottom Decile	661.13	68.65	26.81	622.69	$10,387.91
2nd Decile	810.30	63.20	16.88	763.60	$8,511.83
3rd Decile	706.11	59.61	12.08	676.36	$10,102.81
4th Decile	846.42	51.82	9.86	810.07	$7,859.00
5th Decile	659.28	50.63	10.88	628.73	$7,658.75
6th Decile	921.48	53.80	9.09	879.75	$7,235.00
7th Decile	774.16	48.58	7.60	738.77	$6,766.80
8th Decile	999.81	44.28	12.47	959.80	$6,813.38
9th Decile	1032.04	39.49	16.90	993.92	$7,154.00
Top Decile	837.43	29.38	12.62	806.18	$7,465.04
Total	818.13	51.09	13.53	781.56	$8,039.41

Table 4.16 highlights the relationships between total high school accountability scores and teacher characteristics by dividing accountability scores into deciles. The deciles reveal similar patterns to those found in the elementary school deciles. As average years of teaching experience and percent of teachers with a Masters degree or above increases, the accountability score decile also increases. These findings support investments in furthering teacher education levels and the importance of retaining experienced teachers. They also support the importance of placing such teachers in the schools that serve students with the greatest needs.

Table 4.16

Total High School Accountability Scores in Deciles with Mean Teacher Characteristics

Total Accountability Score Under Model A	Average Years Teaching Experience	Percent of Teachers with a Masters or Above
Bottom Decile	10.63	78.80
2nd Decile	11.53	82.00
3rd Decile	11.04	80.32
4th Decile	11.83	83.88
5th Decile	11.53	78.08
6th Decile	11.66	82.13
7th Decile	11.93	83.16
8th Decile	11.69	80.57
9th Decile	12.44	85.05
Top Decile	12.51	86.20
Total	11.66	81.95

CHAPTER FIVE

CONCLUSIONS AND DISCUSSION

Research Purpose

The purpose of this research was to determine to what extent changes to Kentucky's Accountability Model would result in different conclusions when utilizing the same student data. *No Child Left Behind* and Kentucky's Senate Bill One produce a myriad of rewards and consequences for schools and districts based on their accountability scores. Therefore, it is critical that decisions regarding accountability models be intentional in even the smallest of details. A second purpose of the study was to determine which school types were positively/negatively affected by each of the models presented in the study.

Research Methods

Quantitative methods were used in the study. The researcher used public data supplied by the Kentucky Department of Education. The 2011-2012 accountability data of all 723 elementary and 230 high schools were used. The dependent variable in this study was the total accountability scores that emerged from the 2011-2012 elementary and high school K-PREP test data. The independent variables for this study included three elementary and two high school drafts of the K-PREP accountability model as outlined in tables 3.10 and 3.11. Bivariate correlations were conducted to provide a measure of the strength of the linear associations between the total accountability scores under each model with various school and teacher characteristics in the elementary and

high school accountability models. The top and bottom five percent of schools for each of the models in the study were identified under each model because the rewards and consequences categories for the Unbridled Learning Accountability System revolve around the bottom and top five percent of schools.

Discussion of Research Findings

The research findings revolve around two research questions. Specifically, the conclusions and discussion emerged from the following research questions:

1. To what extent would changes to Kentucky's Unbridled Learning Accountability Model result in different conclusions when utilizing the same student data?
2. Which school types are positively/negatively affected by each of the Kentucky models that were under consideration circa 2010?

Summary of Variations in Results from the Three Elementary Models

Exploring research question number one, to what extent did changes to Kentucky's Accountability Model result in different conclusions when utilizing the same data, as noted in Table 4.1, the relationship of the total score of Model A, the adopted model, with the total score from Model B was an exceptionally high positive correlation, $r(723)=.994$, $p < .001$. The correlation of the total score from $r(723)=.997$, $p < .001$. Finally, a similarly high positive correlation represented the relationship between the total scores from Model B with Model C, $r(723)=.982$, $p < .001$. Collectively, these results indicated that each school would receive a highly comparable total accountability score, regardless of which of the models had been adopted.

Although the total scores were highly correlated across the three models, bivariate correlations were run between the three domains within each model. No major differences were found among the domain correlations with each model. However, these inner correlations were quite telling. In the adopted model, the correlations of Achievement with Gap (r=.88) and Growth (r=.56) were quite high. The implications of these high correlations are discussed below.

The Implications of High Correlations between Elementary Models

The rankings and total accountability scores differed within each of the three models with 30 of the 37 schools identified in Model A also identified within Models B and C. In addition, all of the 37 schools identified in Model A were identified in at least one of the other models. There were only six schools identified in only one of the three models: Cordia, Fulton County, Salyersville, Big Creek, Majestic Knox Center, and Flat Lick. Due to rounding, only 36 schools were identified in the bottom 5% in Models B and C, where 37 school were identified in Model A, the adopted model. Lastly, Models A and B are the most similar with only four schools different between the two models.

Although the correlations between the total scores of three elementary models were statistically significant and only resulted in slightly different results in the bottom five percent rankings of schools for models A, B. and C, these differences have serious consequences for the schools identified. Public perception of the schools is tarnished with the Persistently Low Achieving (PLA) status. Often, the climate and culture of the schools identified become characterized as downtrodden, overwhelmed, and helpless. Fractional differences between the bottom 5% can be the difference between a principal

having a job, a school or district having capacity to lead, and a Site Based Decision Making Council's ability to aid the principal in governing the school. Further, the addition of state Educational Recovery Specialists and Leaders to the identified schools sometimes exacerbates the belief that the teachers and leaders do not know what to do to improve the school.

In 2011, Kentucky identified the bottom five percent of schools and awarded $56 million in School Improvement Grant (SIG) monies to these 41 schools to help improve the educational processes within the schools. With each of the elementary models producing a slightly different bottom five percent, how do we know which list of schools most needs the money and resources provided by the state to improve the educational processes? Additionally, the three elementary models also produced a slightly different top five percent. Although the changes reflected here may seem less drastic than the changes to the bottom five percent, there are still negative consequences to the changes within the top five percent as well. There is a significant amount of positive press at the state and local level when a school or district receives such distinctions. When purchasing or building homes, new homeowners look at the rankings of schools and often want to live in the school districts with the highest performing schools. Enrollment at these schools often increases as a result of the positive identification. These schools often attract the most experienced and educated teachers as well. As a result of this cycle, the rich or highest performing schools get richer and become even higher performers. Unfortunately, the lowest performing schools can get caught in a similar negative cycle.

Another implication of the high correlations between the different model and scores become clear when of the scores of each model were rank ordered. Specifically,

the scores are very close together. In fact, schools were within one point of being in the lowest five percent. This raises the question of the level of precision and reliability of the model. Regardless, just like slightly altering the model would identify a few different schools, slight changes in the cut score would also change the top and bottom percent.

Another critical finding was the high correlation between Achievement with Gap (r=.88) within model A. Such a high correlation could indicate that these two components are essentially the same construct. This strong correlation suggests that two-thirds of the elementary model is an achievement or status score. The data suggests that Gap is simply another achievement score for your non-duplicated Gap groups, as opposed to your whole school achievement score. One of the selling points for educators on the Unbridled Learning Accountability model was that students in the gap groups would only count one time in the accountability model. For example, if a student is non-white and eligible for free and reduced lunch, he will only be counted once, instead of his scores being counted in each of the gap groups he represented. The significantly high correlation between the Achievement score and Gap scores suggests these students are in fact counting twice in the accountability data. According to *NCLB*, these non-duplicated gap groups include historically underperforming non-white students, low income students, students with disabilities, and limited English speaking students. As an example, in a school with 100% of students eligible for free or reduced lunch, the Achievement score would equal the Gap score, because the non-duplicated students are all of the schools' students. Thus, the Gap score does not really measure a gap between traditional lower performing groups of students and their counterparts. The Gap score is, on the contrary, another Achievement score for groups of students that traditionally perform lower, which highlights issues with

the construct and criterion-related validity of the Unbridled Learning Accountability Model.

The data suggest that the adopted accountability model for the elementary schools would disadvantage schools that had served higher percentages of students who do not traditionally perform well, schools with higher percentages of non-white students, students eligible for free and reduced, students with disabilities, and limited English speaking students because the Gap score is not a true Gap score. Rather, it is essentially a second achievement score. The lack of a true Gap score and having a model with two achievement scores not only disadvantages low income schools, it may allow high performing students to mask lower performing groups, at least based only on the total score. It is important to note that Kentucky reports disaggregated data for various student groups and low performance by one or more of these groups results in the school being labeled a Focus School.

Lastly, the high correlation between Achievement and Growth (r=.56), although not as high as, could lead one to contend the Gap is an achievement score as well. *NCLB* mandated that states measure student growth in their accountability systems. Kentucky's answer to this mandate is a quasi-growth model. Student growth is measured over a two year period in comparison to peers scoring at the same point. Regardless of the amount of growth made by students starting at the same baseline, 60% of the students are deemed as making growth, and 40% of the students are identified as not making growth. There is no vertically scaled test given to identify specifically the growth each student made. Students are simply lumped into two categories- made growth or did not make growth, which may or may not be a true picture of the actual growth a student has made. In

reality, all students starting at the same baseline could make one year or more of growth. Despite this hypothetical growth, 40% would be counted in the accountability score as not making growth, which highlights issues with the construct and criterion-related validity of the Unbridled Learning Accountability Model.

The Two High School Models

Exploring question number one, to what extent did changes to Kentucky's Accountability Model result in different conclusions when utilizing the same data from, as noted in Table 4.6, the relationship of the total score of Model A, the adopted model, with Model B was a very high significant positive correlation, $r(230)=.978$, $p< .001$. This result indicated that each school would receive highly comparable total accountability score, regardless of which of the models had been adopted.

Although the total scores were highly correlated between the models, bivariate correlations were run between the five components of each model. No major differences were found among the component correlations, but the specific inner correlations were quite telling. Using Model A, the adopted model, the correlations of Achievement with Gap ($r=.88$) and Growth ($r=.62$) were quite high.

As was the case with the elementary models, one implication of the high correlations between Achievement with Gap ($r=.88$) in the high school model is that these two components are essentially the same score. Like the elementary analyses, this strong correlation suggests that two-fifths of the high school model is an Achievement or status score. The data suggest that Gap is another Achievement score for non-duplicated Gap groups. In addition, the data suggest that the adopted accountability model for the high

schools would also disadvantage schools serving higher percentages of students that do not traditionally perform well, schools with higher percentages of non-white students, students eligible for free and reduced, students with disabilities, and limited English speaking students. This disadvantage is because the Gap score is not a true Gap score. Rather, it is essentially a second Achievement score, which highlights issues with the construct and criterion-related validity of the Unbridled Learning Accountability Model..

Like the elementary model, the high correlation between Achievement and Growth (r=.62), although not as high as Gap, could indicate that Growth is also an achievement score. A second interpretation is that high performing high schools serve higher performing students who also make greater growth than their peers, thus creating underachievement gaps. Regardless, *NCLB* mandated that states measure student growth in their accountability systems. As noted in the section on the elementary model, Kentucky's answer to this mandate is a quasi-growth model. Student growth is measured over a two year period in comparison to peers scoring at the same point the previous year. Regardless of the amount of growth, 60% of the students are deemed as making growth, and 40% of the students are identified as not making growth.

The Implications of High Correlations between High School Models

Although the same eleven schools are in the bottom five percent for both of the high school models, rankings and scores did change as a result of the differences between high school models A and B. With the same eleven schools being identified by both high school models, it is reassuring that the schools that most need the resources to improve instructional resources would be provided them under either model. This triangulation of

data indicates there is reliability in the high school models. Six of the eleven high schools in the bottom five percent for models A and B are in Jefferson County, the largest school district in Kentucky. This finding should lead to rich discussion regarding the fact that more than one-half of the identified schools are from a single district. It is possible that the model itself is a factor.

Table 5.1 highlights the high percentage of free and reduced rates for the bottom five percent of schools for the 2011-2012 school year. The percentage of students eligible for free and reduced lunch rates ranges from 65.8% to 92.4%, and it is widely known that high school students who are eligible often do not apply,. In addition, eight of the eleven bottom five percent high schools also have a percentage of non-white students ranging from 41.3% to 72.9%. Caverna, Cordia, and Phelps High Schools do not have higher percentage rates of non-white students like others in the bottom five percent. However, they all have the lowest total school enrollments on the list. The six bottom five percent schools located in Jefferson County have a percentage of students eligible for free and reduced ranging from 66.7% to 92.4%, and a percentage of non-white students ranging from 40.7% to 72.9%.

Table 5.1

School Characteristics of the High Schools in the Bottom Five Percent

School Name	Total School Enrollment	Percentage of Free and Reduced Students	Percentage of Non-White Students
1. Caverna	196	66.8	19.4
2. Cordia	257	65.8	3.9
3. Doss*	837	71.1	56.9
4. Holmes	707	91.2	41
5. Iroquois*	1033	80.7	67.8
6. Perry Central	903	67.9	1.9
7. Phelps	357	71.7	1.1
8. Southern*	1150	66.7	41.2
9. The Academy at Shawnee*	460	92.4	61.1
10. Valley*	912	71.8	40.7
11. Western*	689	77.3	72.9

* Denotes schools located in Jefferson County

In contrast to the schools in the bottom five percent, nine of the same schools are in the top five percent for both high school models, as noted in table 5.2. However, South Warren and South Oldham are not in the top five percent in model A and were replaced by Paintsville and Hickman in model B. Again, there are still negative consequences to the changes within the top five percent. Unlike the bottom five percent of schools located in Jefferson County, the schools in the top five percent located in Jefferson County have significantly lower percentage of students eligible for free and reduced lunch with ranges from 17.4% to 31.1%, and the percentage of non-white students are also significantly lower with ranges from 31.8% to 41.8%.

Table 5.2

School Characteristics of the High Schools in the Top Five Percent

School Name	Model A Accountability Score	Model B Accountability Score	Mean Total School Enrollment	Mean Percentage of Free and Reduced Students	Mean Percentage of Non-White Students
Beechwood	84.5	85.8	573	12.4	5.9
Brown*	75.7	75.8	713	31.1	41.8
Dupont Manual*	87.3	90.4	1881	17.4	31.8
Hickman		72.8	333	58.8	14.1
Highlands	79.0	78.5	808	14.8	5.6
Louisville Male*	77.0	77.2	1659	23.8	35.6
Model Laboratory	73.5	72.6	208	.50	13.5
Murray	69.6	73.2	438	25.3	16.2
North Oldham	76.0	77.0	999	5.2	7.9
Paintsville		73.1	329	32.8	2.4
South Oldham	68.9		1149	15.5	9.1
South Warren	68.1		897	25.1	10.4
Walton-Verona	77.9	82.0	459	30.9	3.9

*Denotes schools located in Jefferson County.

The Positive/Negative Impact on Types of School from the Elementary Models

Exploring research question two, which elementary school types were positively/negatively affected by each of the three models presented, as noted in Table 4.11, under model A, the adopted model, total student enrollment (r=.18), average years of teaching experience (r=.20), and percentage of teachers with a Masters degree or above (r=.19) were positively related to the total accountability score. In other word,

larger schools that employ teachers who are more experienced and have higher education levels earn higher accountability scores.

In contrast, the relationships between the total accountability score under model A, the adopted model, with percentage of non-white students (r=-.25), percentage of students eligible for free or reduced lunch (r=-.54), and spending per pupil (r=-.28), are negatively related to the accountability score. As these three variables increase, accountability scores decline. Breaking the scores down into deciles further highlights a positive correlation between total accountability scores and teacher characteristics.

The Positive/Negative Impacts on Types of High Schools under Models A and B

The second research question for this study sought to determine which school types were positively/negatively affected by each of the models presented n models A and B, there is statistically significant positive relationship between the total student enrollment, average years of teaching experience, and percent of teachers with a masters or above, and the total accountability score as shown in table 4.12. In other words, as the characteristics increase the accountability scores increase. The top performing high schools had higher total student enrollments, higher average years of teaching experience, and a higher percentage of teachers with a Masters degree or above. The poorest performing schools had lower student enrollments, teacher with less experience, and a lower percentage of teachers with a Masters degree or above.

On the contrary, in models A and B, the adopted model, there is a statistically significant negative relationship between percentage of non-white students, percentage of students eligible for free or reduced lunch, and spending per pupil with the total

accountability scores. In other words, there is a negative impact on the accountability score based on increasing levels of these characteristics. As the percentage of non-white students, the percent of students qualifying for free and reduced lunch, and spending per pupil increase, the accountability score goes down. Both high school models resulted in the same positive and negative relationships between accountability scores and the school characteristics. Therefore, it can be concluded that the high school models would not provide different results.

For example, of all the schools in the top five percent, only one, Hickman, served a student enrollment of which more than 50% were eligible for free or reduced lunch. The school with the lowest free and reduced lunch rate in the bottom five percent was Corida with a 65.8% rate. Clearly, Hickman should be studied with the hope of finding strategies that facilitate the academic success of low income students. One positive note, schools with more experienced and more highly educated teachers produce higher accountability scores for their schools. These data suggest that retaining and educating our most experienced teachers will likely lead to significant achievement gains. Further, the data support the policy that our traditionally lowest performing students should be placed with our most experienced teachers.

Implications for Further Study

The first implication for further research would include researching the formal definitions of gap, growth, and achievement and the alignment of Kentucky's Unbridled Learning Accountability Model and other state accountability models to the formal

definitions as outlined by academic research. This study suggests the alignment of Kentucky's model has low discriminant validity.

Kentucky is currently in the process of a statewide pilot of their new teacher evaluation system, Teacher Professional Growth and Effectiveness System (TPGES), which for the first time in Kentucky's accountability history will include student growth as one of the multiple measures of teacher effectiveness. A correlation between the accountability scores of Kentucky's Unbridled Learning Accountability system with the Teacher Professional Growth and Effectiveness System could provide insight on more teacher and school characteristics and practices about which data should be collected with the hope of identifying variables that can be replicated to increase student achievement, especially in the lowest performing schools. In addition, conducting a study of the effect of leadership behaviors within the bottom and top five percent of schools and their impact on the total score from the u Kentucky's Unbridled Learning Accountability system could provide much needed information. Finally, the data from this study indicate that teacher experience and education are in fact important to student achievement. Therefore, further research is needed on policies and practices that result in attracting and retaining our most experienced and educated teachers.

Summary

Although changes to Kentucky's Accountability Model did not result in widespread differences in conclusions when utilizing the same student data, there were a small number of changes with high stakes consequences for schools. Being identified in the top or bottom five percent of schools can have a major impact on schools. The

different model resulted in a few schools moving into or out of the bottom and top five percent. While the impact was on a small number of schools, the impact would be sweeping and include such measures as replacing principals, terminating school based councils, and placing Educational Recovery Teams in schools. These seemingly small changes can have far-reaching consequences for the schools identified.

Because of the high correlation between the total accountability scores from each of the models presented in this study, all of the models identified the same teacher and school characteristics that positively or negatively influenced accountability scores. All of the elementary and high school models consistently identified positive correlations of the same magnitude between the total school enrollment, average years of teaching experience, and percent of teachers with a Masters or above with total accountability scores. Further, all elementary and high school models also demonstrated comparable negative correlations between the percentage of non-white students, percentage of students eligible for free or reduced lunch, and spending per pupil with total accountability scores. Given the relationships, it is imperative for policymakers to continue to assess the effect of different accountability models on the schools with different student characteristics. Given that the adopted models weight status achievement scores more heavily, regardless of whether they are labeled as such scores in the models, one could argue that these models disadvantage schools that serve higher percentages of traditional lower performing students, low-income, non-white, and students with disabilities. Although the embedded advantages and disadvantages of such models should be discussed given the high stakes resulting from their outcomes, that

discussion should not detract from the relentless pursuit of strategies enabling the academic success of all students, regardless of their backgrounds.

References

-ACT, Inc. (2013). Description of the ACT. Available online: http://www.actstudent.org/testprep/descriptions/

Amrein-Beardsley, A. (2008). Methodical concerns about the education value-added assessment system. *Educational Researcher*, 37 (2) 65-75. doi: 10.3102/0013189X08316420

Amrein-Beardsley, A., Berliner, D.C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education Policy Analysis Archives*, 18 (14). Retrieved from EBSCO*host*.

Ananda, S. & Rabinowitz, S. (2001) Building a workable accountability system. *WestEd*, 1-11.

Auty, W. (2008). Implementer's guide to growth models. *Council of Chief State School Officers*, January, 1-36. Available online: http://www.ccsso.org

Baker, E. (2005). Improving accountability models by using technology-enabled knowledge systems. *Center for the Study of Evaluation*, Report 656.

Bathgate, K., Colvin, R., & Silva, E. (2011). Striving for student success: A model of shared accountability. *Education Sector Reports: Striving for Student Success*, November. Available online: http://www.educationsector.org

Betebenner, D. & Linn, R.L. (2009). Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability. *Educational Testing Service*, December, 3-23.

Brookhart, S.M. (2009). The many meaning of "Multiple Measures." *Educational Leadership: Multiple Measures*, 67 (3), 6-12.

Carey, K. &Manwaring, R. (2011). Growth models and accountability: A recipe for remaking ESEA. *Education Sector Reports: Growth Models and Accountability*, May. Available online: http://www.educationsector.org

Chester, M.D. (2005). Making valid and consistent inferences about school effectiveness from multiple measures. *Educational Measurement: Issues and Practice*, 24 (4), 40-52.

Choi, K., Goldschmidt, P. &Yamashiro, K. (2006), Exploring models of school performance: From theory to practice. *Yearbook of the National Society for the Study of Education*, 104: 119–146.

Dobb, R., Madgavkar, A., Burton, D., Labaye, E., Manyika, J., Roxburgh, C., Lund, S., & Madhav, S. (2012). The world at work: Jobs, pay, and skills for 3.5 billion people. McKinsey Global Institute. Available online: www.mckinsey.com/mgi

Draut, K. (2011). Comparison of NCLB and Kentucky's accountability model. *Kentucky Department of Education*, Office of Assessment and Accountability, August 5.

Draut, K. (personal communication, 2013, January 9). Re: Information request.

Finnigan, K.S. & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal*, 44 (3), Retrieved from EBSCO*host*.

Foster, J. (1991). The role of accountability in Kentucky's education reform act of 1990. *Educational Leadership*. ASCD. 48 (5), 34-36.

Fuller, B., Wright, J. Gesicki, K. & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher*, 36(5), 268-278. Doi: 10.3102/0013189X07306556 Available online: http://edr.sagepub.com/content/36/5/268

Gabriel,T. (2010). Cheat sheet: Under pressure, teachers tamper with tests. *New York Times*. June.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, Wi., Hebbler, S., Blank, R. & Williams, A. (2005). Policymakers' guide to growth models for school accountability: How do accountability models differ? *Council of Chief State School Officers*, October, 1-27. Available online: http://www.ccsso.org

Goldstein, J. & Behuniak, P. (2005) Growth models in action: Selected case studies. *Practical Assessment Research & Evaluation*. 10 (11). Available online: http://pareonline.net/getyn.asp?v=10&n=11

Greenhouse, M., Harris, M., Li, K., Looney, A. & Patashnik, J. (2012). A dozen economic facts about k-12 education. The Hamilton Project, Washington, D.C. Available online: www.hamiltonproject.org

Gross, L. (2012). First results from unbridled learning accountability model released. Kentucky Department of Education. Available online: http://www.education.ky.gov

Gross, L. (2009). News releases on assessment, accountability and standards. Kentucky Department of Education. 1-47.

Haertel, E. & Herman, J. (2006). A historical perspective on validity arguments for accountability testing. *National Center for Research on Evaluation, Standards, and Student Testing*, CSE Report 654.

Herman, J. (2007). Accountability and assessment: Is public interest in k-12 education being served? *National Center for Research on Evaluation, Standards, and Student Testing* (CRESST). Report 728.

Herman, J. & Dietel, R. (2005). A primer on accountability. *American School Board Journal,* December 2005.

Herrera, S. & Murry, K.G. (2006). Accountability by assumption: Implications of reform agendas for teacher preparation. *Journal of Latinos and Education*, 5 (3), 189-207. Retrieved by EBSCO*host*.

Ho, A.D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, 37 (6), 351-360. doi:10.3102/0012189X083223842
Available online: http://er.aera.net

Jones, B.D., &Egley, R.J. (2007). Learning to take tests or learning for understanding? Teachers' beliefs about test-based accountability. *The Educational Forum*, 71 (3), 232-248. Retrieved from EBSCO*host*.

Kentucky Board of Education. (2010). Goals and guiding principles for accountability in Kentucky's public education system. Available online: http://dspace.kdla.ky.gov:8080/xmlui/handle/10602/11617

Kentucky Department of Education. (2011A). Superintendant's annual attendance report. Available online: www.education.ky.gov

Kentucky Department of Education. (2011B). 2011 K-PREP number of items and testing times. Available online: http://education.ky.gov/AA/distsupp/Pages/K-PREP.aspx

Kentucky Department of Education. (2011C, January). Next generation learners proposed accountability model: Draft for discussion. Office of Accountability.

Kentucky Department of Education. (2011D, April). Next generation learners proposed accountability model: Draft for discussion. Office of Accountability.

Kentucky Department of Education. (2011E, July). Next generation learners proposed accountability model: Draft for discussion. Office of Accountability.

Kentucky Department of Education. (2012A). Kentucky school report card. Available online: http://applications.education.ky.gov/src/ProfileByState.aspx

Kentucky Department of Education. (2012B, June). Next generation learners proposed accountability model: Draft for discussion. Office of Accountability.

Kentucky Department of Education. (2013). Kentucky's high school end-of-course assessments: Answers to parents' most frequently asked questions. Available online: http://education.ky.gov/AA/distsupp/Pages/EOC.aspx

Kim, J.S. & Sunderman, G.L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34 (3). 3-13. doi:10.3102/0013189X034008003 Available online: http://edr.sagepub.com/content/34/8/3

Kress, S., Zechmann, S., & Schmitten, J. (2011). When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Journal on Legislation*.

Lauermann, F. & Karabenick, S.A. (2011). Taking teacher responsibility into account(ability): Explicating multiple components and theoretical status. *Educational Psychologist*, 46 (2), 122-140. Doi:10.1080/00461520.2011.558818

Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, 78 (3), 608-644. Doi: 10.3102/0034654308324427 Available online: http://rer.aera.net

Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*, 29 (4), 4-16. Doi:10.3102/0013189X029002004 Available online: http://edr.sagepub.com/content/29/2/4

Linn, R.L. (2003). Accountability: Responsibility and reasonable expectations. Educational Researcher, 32 (7), 3-13. Doi:10.3102/0012189X032007003 Available online: http://edr.sagepub.com/content/32/7/3

Linn, R.L. (2003). Performance standards: Utility for different uses of assessments. Educational Policy Analysis Archives, 11 (31). Retrieved October 26, 2012 from http://epaa.asu.edu/epaa.v11n31

Linn, R.L. (2005A). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. Educational Policy Analysis Archives, 13 (33), 1-20.

Linn, R.L. (2005B). Issues in the design of accountability systems. *National Center for Research on Evaluation*, Technical Report 650.

Linn, R.L. (2005C). Test-based educational accountability in the era of No Child Left Behind. CSE Report 651. National Center for Research on Evaluation, Standards, and Student Testing.

Linn, R.L. (2006). Educational accountability systems. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Marcus, J. (2012). Opt-out movement gains steam. *Harvard Education Letter*, 28 (2).

Martel, E. (2011). The Atlanta Scandal: Teaching in "A Culture of Fear, Intimidation and Retaliation." *The Washington Post*.

Martineau, J. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics,* 31 (1), 35-62. doi:10.3102/10769986031001035

Maruyama, G. (2012). Assessing college readiness: Should we be satisfied with ACT or other threshold scores? *Educational Researcher*, 41 (7), 252-261. doi:10.3102/0013189X12455095

Massell, D. & Fuhrman, S. (1994). Ten years of state education reform, 1983-1993: Overview with four case studies. Center for Policy Research in Education. 1-170.

McEachin, A. & Polikoff, M. (2012). We are the 5%: Which schools would be held accountable under a proposed revision of the Elementary and Secondary Act? *Educational Researcher* 41 (7), 243-251. doi: 10.3102/0013189X12453494

Meier, D. &Wood, G. (Eds.). (2004). Many children left behind: How the No Child Left Behind Act is damaging our children and our schools. Boston: Beacon Press.

Office of Assessment and Accountability. (2012). 2012 K-Prep number of items and testing items. Kentucky Department of Education. Available online: http://education.ky.gov/AA/distsupp/Documents/2012KPREPTestingFormatItemsandTimes20120312.pdf

O'Malley, K., Murphy, S., McClarty, K., Murphy, D., & McBride, Y. (2011). Overview of student growth models. *Pearson*.

Mintrop, H. & Sunderman, L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement—and why we may retain it anyway. Educational Researcher 38 (5), 353-364. Available online: edr.sagepub.com/content/38/5/353

No Child Left Behind Act of 2001, Pub. L. 107-110. Available online: www.ed.gov/policy/elsec/leg.html

Perie, M., Park, J. & Klau, K. (2007). Key elements for educational accountability models. Council of Chief State School Officers.

Popham, W.J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Quinn, S., & Ethridge, E. (2006). Education in an era of accountability: Do you have to sacrifice wise practices? *Early Childhood Education Journal*, 34 (2), 117-123. doi:10.1007/s10643-006-0066-2

Rabibowitz, S., Roeber, E., Schroeder, C., & Sheinker, J. (2006). Creating aligned standards and assessment systems. The Council of Chief State School Officers.

Ready, D. (2012). Associations between student achievement and student learning: Implications for value-added school accountability models. *Educational Policy*, May, 1-29. Doi: 10.1177/0895904811429289

Reardon, R. (2011). Percentile growth modeling: A policy response to educational accountability. Metropolitan Educational Research Consortium (MERC), November.

Schafer, W.D. (2003). A state perspective on multiple measures and school accountability. *Educational Measurement: Issues and Practice*, 22 (3), 27-31.

Scot, T., Callahan., C.M., & Urquhart, J. (2009). Paint-by-number teachers and cookie-cutter students:The unintended effects of high-stakes testing on the education of gifted students. *Roeper Review*, 31 (1), 40-52. Doi:10.1080/02783190802527364

Shepard, L. (1980). Technical issues in minimum competency testing. *Review of Research in Education*, 8, 30-82. *American Educational Research Association*. http://www.jstor.org/stable/1167123

Seiler, M., Lunney, G., Olds, S., & Young P. (2005). An analysis of the commonwealth accountability testing system. Legislative Research Committee. Available Online: http://www.lrc.ky.gov/lrcpubs/RR328.pdf

Sims, R. (2012, August). Personal Interview. Kentucky Department of Education: Educational Recovery Assessment Update. Lexington, KY.

Sims, R. (2013). Kentucky testing system. Available Online: http://education.ky.gov/AA/KTS/Pages/default.aspx

Stone, C.A., & Lane, S. (2003). Consequences of a state accountability program: Examining relationships between school performance gains and teacher, student, and school variables. *Applied Measurement in Education*, 16 (1). Retrieved from EBSCO*host*.

Tankersley, K. (2007). Tests that teach: Using standardized tests to improve instruction. Alexandria, VA: Association for Supervision and Curriculum Development.

Thum, Y.M. (2003). No Child Left Behind: Methodogical challenges & recommendations for measuring adequate yearly progress. *Center for the Study of Evaluation*, Tech Report 590.

Tungate, M. (2010). New accountability model begins to take shape. *Kentucky Teacher*, December. Available online: http://www.kentuckyteacher.org/news/2010/12/new-accountability-model-begins- to-take-shape/

Ujifusa, A. (2012). Scores drop on Ky.'s common core- aligned test. Education Week, November 4. Available online: http://www.edweek.org/ew/articles/2012/11/02/11standards.h32.html?intc=es

Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45 (2), 107-122. Retrieved from EBSCO*host*.

Yu, F., Kennedy, E., Teddlie, C., &Crain, M. (2007). Identifying effective and ineffective schools for accountability purposes: A comparison of four generic types of accountability models. *Council of Chief State School Officers*, 1-53.

Zvoch, K. & Stevens, J. (2008). Measuring and evaluating school performance: An investigation of status and growth-based achievement indicators. *Evaluation Review*, 32 (6), 569-595. Doi: 10.1177/0193841X08320398 Available online: http://erx.sagepub.com/content/32/6/569

Appendix: Next- Generation Learners Proposed Accountability Model

NEXT-GENERATION LEARNERS PROPOSED ACCOUNTABILITY MODEL

DRAFT FOR DISCUSSION

Revised based on input from Kentucky Board of Education, stakeholders and data simulations

The Kentucky Department of Education's mission is to prepare all Kentucky students for next-generation learning, work and citizenship by engaging schools, districts, families and communities through excellent leadership, service and support.

BACKGROUND

Education Commissioner Terry Holliday and staff in the Kentucky Department of Education continue to discuss with the Kentucky Board of Education (KBE) and various stakeholder groups (i.e., School Curriculum, Assessment and Accountability Council (SCAAC), Superintendents in Co-op meetings, District Assessment Coordinators, Kentucky Association of Assessment Coordinators, Education Coalition, Math Achievement Committee, Kentucky Association of School Councils Conference, Prichard Committee for Academic Excellence and Parents Advisory Council) the broad concepts proposed for a future state accountability model. Specifically, the broad categories of Achievement, Gap, Growth, Readiness and Graduation Rate are being introduced to solicit feedback from educators, stakeholders and the public.

On December 7, 2010, the Kentucky Board of Education (KBE) participated in a study session regarding the proposed accountability model. The study session yielded several KBE decisions that are reflected in this document. Based on stakeholder feedback and data simulations, revisions have been made to simplify the data calculations and reduce complexity.

A BALANCED APPROACH

Senate Bill 1 (2009 Kentucky General Assembly) requires Kentucky to begin a new assessment and accountability system in 2011-2012. The proposed assessment and accountability model is a balanced approach that incorporates all aspects of school and district work and is organized around the Kentucky Board of Education's four strategic priorities: next-generation learners, next-generation professionals, next-generation support systems and next-generation schools/districts.

The list below details the indicators that could be included in the future accountability model around each of these strategic priorities.

Next-Generation Learners	Next-Generation Professionals	Next-Generation Support Systems	Next-Generation Schools/Districts
Achievement (Proficiency) Gap Growth Readiness for College/Career Graduation Rate	Percent Effective Teachers Percent Effective Leaders	Working Conditions Survey Program Reviews	Revised Report Card New Accountability System

The attached document is an overview of the proposed accountability model for next-generation learners.

NEXT-GENERATION LEARNERS PROPOSED ACCOUNTABILITY MODEL

DRAFT FOR DISCUSSION

Revised based on input from Kentucky Board of Education, stakeholders and data simulations

Calculation for School/District Point Total

Points generated in **Achievement** for all 5 content areas **+ Gap** (percentage of proficient and distinguished) for the Non-duplicated Student Group for all 5 content areas **+ Growth** in reading and mathematics (percentage of students at typical or higher levels of growth) **+ College Readiness** as measured by the percentage of students meeting benchmarks in 3 content areas on EXPLORE at middle school **+ College/Career Readiness Rate** as measured by ACT benchmarks, college placement tests and career measures **+ Graduation Rate.**

KBE asked that within each Classification an indicator be added to show the direction in which the performance of the school/district is moving.

SCHOOL AND DISTRICT CLASSIFICATIONS

Distinguished	Cut score (to be determined) points or more in **Elementary: Achievement + Gap + Growth** **Middle: Achievement + Gap + Growth + College Readiness** **High: Achievement + Gap + Growth+ College/Career Readiness Rate + Graduation Rate**
Proficient	Cut score (to be determined) points in **Elementary: Achievement + Gap + Growth** **Middle: Achievement + Gap + Growth + College Readiness** **High: Achievement + Gap + Growth+ College/Career Readiness Rate + Graduation Rate**
Needs Improvement	Cut score (to be determined) points in **Elementary: Achievement + Gap + Growth** **Middle: Achievement + Gap + Growth + College Readiness** **High: Achievement + Gap + Growth+ College/Career Readiness Rate + Graduation Rate**
Persistently Low Achieving	Fewer than cut score (to be determined) points in **Elementary: Achievement + Gap + Growth** **Middle: Achievement + Gap + Growth + College Readiness** **High: Achievement + Gap + Growth+ College/Career Readiness Rate + Graduation Rate**

NEXT-GENERATION LEARNERS PROPOSED ACCOUNTABILITY MODEL

DRAFT FOR DISCUSSION

Revised based on input from Kentucky Board of Education, stakeholders and data simulations

Categories within Next-Generation Learners

(This model is based on student data from state-required assessments administered in grades 3-12.)

Grade Range	Achievement	Gap	Growth	College/Career Readiness	Graduation Rate
Elementary	Tests: Reading, mathematics, science, social studies and writing	Tests: Reading, mathematics, science, social studies and writing	Reading and mathematics	N/A	N/A
Middle	Tests: Reading, mathematics, science, social studies and writing	Tests: Reading, mathematics, science, social studies and writing	Reading and mathematics	EXPLORE (College Readiness)	N/A
High	End of Course Tests**	End of Course Tests**	PLAN to ACT *Reading and mathematics*	College/Career Readiness Rate	AFGR*/Cohort Model

*AFGR is Averaged Freshman Graduation Rate.

**SCAAC has recommended four End of Course exams in 2012, the first year of the new system: English II, Algebra II, Biology and US History.

Process

Individual student data collected from the assessments and rates listed in the chart above are used to generate a numeric value for each category of Next-Generation Learners—Achievement, Gap, Growth, College Readiness and Graduation Rate. The value for each category is weighted to create a final overall score for Next-Generation Learners. The following table illustrates the weights.

Grade Range	Achievement	Gap	Growth	College Readiness	Graduation Rate	Total
Elementary	30	30	40	N/A	N/A	100
Middle	28	28	28	16	N/A	100
High	20	20	20	20	20	100

A standard setting process will establish the cut scores to classify a school or district as Distinguished, Proficient, Needs Improvement or Persistently Low Achieving (PLA). A cut score is the numeric values where schools or districts enter or exit the classifications. Note: The PLA designation identifies the lowest five percent as required by federal and state statute and regulation.