

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

8-2016

Curriculum, classroom assessment, instruction, and science outcomes : evidence from Kentucky's scholastic audits.

Steven Patrick Riley
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

Part of the [Curriculum and Instruction Commons](#)

Recommended Citation

Riley, Steven Patrick, "Curriculum, classroom assessment, instruction, and science outcomes : evidence from Kentucky's scholastic audits." (2016). *Electronic Theses and Dissertations*. Paper 2516.
<https://doi.org/10.18297/etd/2516>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

CURRICULUM, CLASSROOM ASSESSMENT, INSTRUCTION, AND SCIENCE
OUTCOMES: EVIDENCE FROM KENTUCKY'S SCHOLASTIC AUDITS

By

Steven Patrick Riley
B.A., Centre College
M.A.T., Oakland City University

A Dissertation
Submitted to the Faculty of the
College of Education and Human Development of the University of Louisville
and
Graduate Studies and Research at Western Kentucky University
In Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in
Educational Leadership and Organizational Development

Department of Leadership, Foundations, and Human Resource Education
University of Louisville
and
College of Education and Behavioral Sciences
Western Kentucky University

August 2016

CURRICULUM, CLASSROOM ASSESSMENT, INSTRUCTION, AND SCIENCE
OUTCOMES: EVIDENCE FROM KENTUCKY'S SCHOLASTIC AUDITS

By

Steven Patrick Riley
B.A., Centre College
M.A.T., Oakland City University

A Dissertation Approved on

July 29, 2016

by the following Dissertation Committee:

Dr. Stephen Miller
Dissertation Chair

Dr. Douglas Smith
Methodologist

Dr. Thomas Tretter

Dr. Bradley Carpenter

DEDICATION

This dissertation is dedicated to the spiritual, professional, and personal counselors
whose guidance have forever impacted my life

God
(my Spiritual Father)

“I can do all things through Christ who strengthens me.” (Philippians 4:13);

Dr. Vicki Riley
(my mother)

who was there from the beginning—the good times and the bad—and forever serves as my
touchstone for professionalism, my model of sublime empathy, and my cardinal of hope.
“All that I am or ever hope to be, I owe to my angel mother.” (Abraham Lincoln);

Mr. and Mrs. Charles and Freida Ross
(my grandparents)

who offered sage wisdom and bolstered me with unending support;

Miss Alexandria Violet Dulay and Master Philip Dulay-Swickard
(my partner and her son)

who afforded me daily affirmation, encouragement, and love in the final year of my
academic journey and who can now believe that anything is possible;

Mr. Steve Riley
(my father)

who can now beat his chest and yell from the rafters, “My son is a doctor!”

ACKNOWLEDGEMENTS

Dr. Stephen Miller, Western Kentucky University, dissertation chair, your dedication to the craft of the doctoral experience and your unmatched scholarship has not gone unnoticed over the past nine years. You have taught me discipline—both professionally and personally—and patience through the art of dissertation writing. Your kindness and empathy throughout my journey have indeed inspired me—not only to finish my dissertation, but to be a better man. Sir Isaac Newton once said, “If I have seen further than others, it is by standing upon the shoulders of giants.” Dr. Miller, you are my intellectual and academic giant.

Dr. Douglas Smith, Western Kentucky University, methodologist, your support during the results and analysis portion of my dissertation was a blessing. You guided me through SPSS and Chapter IV with patience and encouragement. Thank you for all that you did to make this dissertation a reality, including serving as my IRB representative for Western Kentucky University.

Dr. Thomas Tretter, University of Louisville, dissertation committee member, your expertise in science pedagogy and research methodology truly added value to my dissertation. Additionally, thank you for your willingness to serve as my IRB liaison for the University of Louisville. Dr. Bradley Carpenter, University of Louisville, dissertation committee member, your thoughtful feedback contributed to the overall quality of my dissertation. Thank you for participating on my dissertation committee.

Thank you to the following individuals at Western Kentucky University and the

University of Louisville: Kelly Ising, UofL; Carolyn Hunt, WKU; Courtney Kerr, UofL;
Dr. Bud Schlinker, WKU; Dr. Ric Keaster, WKU; and, Roxanne Spencer, WKU.

ABSTRACT

ACADEMIC PERFORMANCE, ACCOUNTABILITY, AND SCIENCE OUTCOMES: EVIDENCE FROM KENTUCKY'S SCHOLASTIC AUDITS

Steven Patrick Riley

July 30, 2016

Through the Kentucky Education Reform Act of 1990, Kentucky enacted sweeping reform that sought to equalize educational finance. With the foundational principle that all children can learn at high levels, the Kentucky Department of Education penned the Standards and Indicators for School Improvement (SISI) to guide the Scholastic Audit—a school improvement process, underscored by the nine standards and 88 indicators of the SISI.

With the theoretical framework of Systems Theory, the current research investigates Kentucky reform through the lens of the SISI and Scholastic Audit. The central research question for the current study is the following: To what extent is Academic Performance (Standards 1-3 from Kentucky's SISI) related to elementary science accountability scores?

The current study employs a secondary Scholastic Audit database maintained by the KDE of 181 elementary schools ($N = 181$) who conducted a Scholastic Audit between 2001-2005. Additionally, repeat incidences of Scholastic Audit cases were removed to create a purposeful sample of 164 elementary schools ($n = 164$).

Descriptive statistics were performed to assess various possible relationships that could affect school achievement outcomes. Factor Analysis and Cronbach's (1951) coefficient alpha were employed to assess the possibility of data reduction for the Academic Performance standards and their respective indicators. Simultaneous multiple regressions were used to calculate possible relationships among the demographic factors, the Academic Performance standards, and the Science Academic Index. Additionally, a hierarchical multiple regression equation was employed to determine the relationships of the three Academic Performance standards on the Science Academic Index, while controlling for demographic factors.

Standards 1-3 collectively explained 28% of the variance on the elementary science portion of the Academic Index, while only Standard 2, Classroom Evaluation/Assessment yielded a significant impact. Demographic factors accounted for 62.6% of the overall variance on the elementary science component of the Academic Index. The three Academic Performance standards remained fairly independent of demographic factor influence. During the hierarchical multiple regression equation, the addition of the three SISI Academic Performance standards in Step 2 added .046 in Adjusted R² value. The final hierarchical regression model explained 68.0% of the total variance on the Science Academic Index.

TABLE OF CONTENTS

	PAGE
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
LIST OF TABLES	xi
LIST OF FIGURES	xiii
CHAPTER	
I. STATEMENT OF THE PROBLEM.....	1
Introduction.....	1
Elementary Science.....	2
Accountability.....	4
The Problem Defined	10
Purpose.....	13
Research Questions	15
Significance of the Study	18
Limitations of the Study.....	19
Summary.....	21
II. REVIEW OF THE LITERATURE.....	24
Introduction.....	24

Accountability.....	25
Academic Performance	55
Elementary Science.....	76
Academic Performance, Accountability, and Science Outcomes.....	120
Summary	148
III. METHODOLOGY.....	154
Introduction.....	154
Research Design.....	154
Definition of Terms.....	155
Description of the Data	162
The Accountability Process	163
Scholastic Audit Methodology	165
Description of the Variables	167
Analysis Plan.....	175
Empirical Research Questions	181
Reliability and Validity.....	184
Ethical Issues.....	188
Summary	188
IV. RESULTS	191
Introduction.....	191
Descriptive Statistics.....	193
Psychometric Analysis	201
Research Questions	221
Summary	234

V. DISCUSSION AND CONCLUSIONS	239
Introduction.....	239
The Study in Brief.....	244
Discussion	248
Recommendations	276
Conclusions	292
ENDNOTES.....	302
REFERENCES	303
APPENDIX A.....	323
APPENDIX B	326
CURRICULUM VITAE.....	329

LIST OF TABLES

TABLE	PAGE
1. Variables and Type of Multiple Regression by Research Questions	183
2. Descriptive Statistics for Demographic Controls ($N = 164$)	197
3. Descriptive Statistics for Demographic Controls on Voluntary Scholastic Audit Schools ($N = 130$)	199
4. Descriptive Statistics for Demographic Controls on Forced Scholastic Audit Schools ($N = 34$)	200
5. Total Variance Explained for Standard 1, Curriculum	205
6. Factor Loadings for Standard 1, Curriculum	207
7. Internal Reliability and Item Characteristics for Standard 1, Curriculum ($N = 164$)	208
8. Total Variance Explained for Standard 2, Classroom Evaluation/Assessment	209
9. Factor Loadings for Standard 2, Classroom Evaluation/Assessment	211
10. Internal Reliability and Item Characteristics for Standard 2, Classroom Evaluation/Assessment ($N = 164$)	212
11. Total Variance Explained for Standard 3, Instruction	213
12. Factor Loadings for Standard 3, Instruction	215
13. Internal Reliability and Item Characteristics for Standard 3, Instruction	216

14. Correlations for Demographic Factors, Curriculum (Standard 1), Classroom Evaluation/Assessment (Standard 2), Instruction (Standard 3), and Academic Index ($N = 164$).....	219
15. Regression of Standard 1, Curriculum, on the Demographic Variables ($N = 164$).....	224
16. Regression of Standard 2, Classroom Evaluation/Assessment, on the Demographic Variables ($N = 164$)	225
17. Regression of Standard 3, Instruction, on the Demographic Variables ($N = 164$).....	226
18. Regression of the Academic Index (Elementary Science) on the Demographic Variables ($N = 164$).....	228
19. Regression of the Academic Index (Elementary Science) on the SISI Academic Performance Standards (Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction) ($N = 164$)	229
20. Regression of the Academic Index (Elementary Science) on the SISI Academic Performance Standards (Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction), Controlling for Demographic Factors ($N = 164$).....	233

LIST OF FIGURES

FIGURE	PAGE
1. Hypothesized relationships among Demographic Factors, Academic Performance (Curriculum; Classroom Evaluation and Assessment; Instruction), and Science Accountability Outcomes	17
2. Scree plot for Standard 1, Curriculum	206
3. Scree plot for Standard 2, Classroom Evaluation/Assessment	210
4. Scree plot for Standard 3, Instruction	214

CHAPTER I
STATEMENT OF THE PROBLEM

Introduction

The global professional landscape evolved exponentially during the first part of the 21st century—in both type of occupational opportunity and availability. Where the tools of the industrial trade remained arguably constant for nearly a century, and where America subsequently achieved success through the ingenuity and resilience of scientists, engineers, and mathematicians, no longer can America use the same educational pathways and expect similar results. During the 20th century, students received a compulsory education focused on the types of jobs offered in the 20th century. However, this type of kindergarten through twelfth grade curricular prescription will not meet the increasingly rigorous demands for critical thinking and creativity for the science, technology, engineering, and mathematics (STEM) jobs of the 21st century—with many of these jobs not even invented yet. Roberts (2012) cites that “innovation and invention are influential forces in the economy” and that “these skills cannot be developed apart from an education system that prepares students to fulfill these demands” (p. 2). In 2012, the National Research Council (2010) purported that the United States, among 24 industrialized nations, ranked 18th in graduation rates. The National Research Council (2010) also alarmingly reported that while the United States led the world in high school diplomas and college degrees twenty years ago, the United States ranked ninth in the percentage of population with high school

degrees and seventh for college degrees in 2012. Perhaps even more disconcerting is the low number of students pursuing expertise in these content areas—and from educators not versed in the content and skilled in the pedagogy of science, technology, engineering, and mathematics. With only 16% of high school seniors interested in pursuing an education in a STEM field, and only half of those students prepared in a STEM field actually accepting a career in STEM, America is currently ill-equipped to meet the shifting job market (U.S. Department of Education, 2015). Among industrialized nations, America now ranks 25th internationally in mathematics and 17th in science (U.S. Department of Education, 2015).

STEM proponents in government and education endorse this push toward the infusion of science, technology, engineering, and mathematics in current curriculum as students will subsequently be better equipped to meet the demand of professional opportunities in STEM fields (Brown, Brown, Reardon, & Merrill, 2011, p. 5). According to Havice (2009), it is the responsibility of schools to develop 21st century learners and thinkers, with STEM as the underpinning. Roberts (2012) postulates that “STEM education is becoming the logical tool to impart these skills at the secondary level, and even at the primary level” (p. 3).

Elementary Science

In the world of education, the acronym STEM (Science, Technology, Engineering, and Math) is used frequently as a learner goal for integration and enhancement, in direct and indirect initiatives to bring science to the forefront of the teaching and learning landscape for students. While many middle school and high school teachers receive focused content preparation for teaching science, this researcher has noted over the past 10 years that elementary teachers lack the same subject-specific training received by their secondary counterparts.

In 2000, the National Science Foundation (NSF) prepared a research document titled the *2000 National Survey of Science and Mathematics Education*. The survey sought to provide “up-to-date information and to identify trends in the areas of teacher background and experience, curriculum and instruction, and the availability and use of instructional resources” (Fulp, 2002, p. 1). With a response rate of 74%, 5,728 science and mathematics teachers in schools in the United States participated in the survey. In the study, the NSF researchers analyzed the proficiency of science and mathematics teachers, in terms of content and pedagogy. Also, researchers examined what teachers were trying to teach in mathematics and science, and the resources used to achieve these learner goals (Fulp, 2002, p.1). In the sub-report developed by Fulp (2002) for the analysis of elementary (grades kindergarten through 5th grade) school science instruction, the researcher investigated trends and patterns for the 655 science teachers (320 grades K-2 teachers and 335 grades 3-5 teachers). In terms of general demographics, Fulp determined that science teachers in this survey were predominantly white females. Also, 42% had a master’s degree and 60% of the examined population were over the age of 40 (Fulp, 2002, p. 2).

Under the heading “content preparedness,” Fulp (2002, p. 2) cited national standards for urging science instruction to begin in the early grades, with the assumption that science teachers must therefore exhibit competency in the science content. However, according to this survey, only four percent of elementary school science teachers had undergraduate degrees in science or science education—with 86% possessing majors in education. Of those 86% without science or science education backgrounds, 40% had taken four or fewer semesters of college science classes (Fulp, 2002, pp. 2-3). Even in the reporting of perceptions of qualifications for teaching science subjects in elementary

school, a large majority of teachers purported a “not well qualified” or “adequately qualified” rating, with only 28% reporting “very well qualified” in Life Science, 24% reporting “very well qualified” in Earth Science, and 14% reporting “very well qualified” in Physical Science—a stark contrast to the 66% in Mathematics and 77% in Reading/Language Arts who reported being “very well qualified” (p. 5).

In the summary of findings, Fulp (2002) contrasted the high degree with which elementary school science teachers reported high pedagogical preparedness in general methods with the low degree of practices frequently connected with science standards—e.g., developing conceptual understanding of students, making connections between science and other disciplines, and teaching “investigative strategies” (p. 19). In addition, Fulp reported the need by elementary school science teachers for content specific professional development for instructional technology and personal science content knowledge, both assumed to enhance student achievement on science learner outcomes (p. 19).

Darling-Hammond (2007) gave a stark prediction about the future of the United States, in relationship to quality teaching: “The notion that we can remain a world-class economy while undereducating large portions of our population—in particular, students of color and new immigrants, who are fast becoming a majority in our public schools—is untenable” (p. 42). As the volume of STEM in education has steadily increased over the past five years, along with the demand for better prepared elementary school science teachers over the past ten years, more research and investigation is warranted to determine the effects of curriculum, classroom evaluation and assessment, and instruction on elementary science achievement outcomes.

Accountability

At the national level, the *No Child Left Behind Act of 2001* (NCLB), a reform-oriented extension of the *Elementary and Secondary Education Act of 1965* (ESEA), served as a guiding document for the onset of the 21st century. According to the document *No Child Left Behind: A Toolkit for Teachers* (U.S. Department of Education, 2004), NCLB focused on four main reform measures to enhance student achievement: (a) accountability for results, (b) research-based instructional strategies, (c) greater parental options, and (d) more local governance for education (p. 1). Two of the four founding principles of the NCLB document focus squarely on the primacy of sound instruction—the use of research-based instructional strategies and the role of data from accountability assessment results to drive future teaching. In terms of NCLB minimum standards for meeting the Highly Qualified Teacher provision, a teacher must possess a bachelor degree, full state certification or licensure, and demonstrate competency in the subject taught (U.S. Department of Education, 2004).

At the state level, Kentucky devised the *Kentucky Education Reform Act of 1990* (KERA) in a response to the call for equity and the improvement of educational structures of the 1980s, and specifically, *A Nation at Risk* (National Commission on Excellence in Education, 1983). Sixty-six Kentucky school districts filed a law suit with the Kentucky Supreme Court, which sided with the school districts, in declaring the current Kentucky educational program unconstitutional. The lawsuit precipitated state legislators to develop KERA for the reformation of educational practices, with a focus on governance, finance equity, and educational accountability for Kentucky.

Locally, many Kentucky school districts are involved in communities of practice and guided educational cooperative initiatives whereby individual teachers from a participating school join a network of like-minded teachers from other schools to engage in

professional learning and reflection to improve upon their practice. Examples of such professional learning communities are the following: Learning by Design Collaborative in Daviess County Public Schools, the District of Innovation Grant in technology for Owensboro Public Schools, the District of Innovation Grant in College and Career Readiness for Taylor County Public Schools, and the Race to the Top Grant in data analysis, leadership, standards-based grading, and personalized learning for the Green River Regional Educational Cooperative (GRREC) and Ohio Valley Educational Cooperative (OVEC)—affecting 113 elementary and secondary Kentucky schools in 22 school districts.

The Kentucky Education Reform Act

In 1990, the Kentucky Supreme Court, in response to a lawsuit brought forth by 66 Kentucky school districts over flaws in education funding, ruled that Kentucky's education was both "inequitable and inefficient" (Reedy, 2010). From this ruling, Kentucky legislators spearheaded an educational reform movement, the most sweeping of its time, to improve finance, governance, and curriculum (Steffy, 1993; Reedy, 2010). For the first time in Kentucky history, legislators devised measures to ensure a consistent, high quality education, regardless of socioeconomic background of either the student or school district. Different from other reform movements of the 1980s and 1990s, the *Kentucky Education Reform Act of 1990* (KERA) sought to focus not just on setting higher standards, but also examine and improve organizational structures and statewide assessment procedures (Reedy, 2010).

The idea that schools teach at high levels and must ensure all children learn at high levels, in order to effect value-added improvement, underscored KERA (cf. Miller, 1992). According to the Kentucky Department of Education (KDE, 2000b), Kentucky would (a)

provide equitable resources to all districts and schools, (b) provide extra resources for districts and schools with many disadvantaged children, (c) eliminate the political obstacles that impede learning, (d) set high standards of performance for all vested stakeholders, (e) establish structures for technology and communication for teacher and student support, (f) enable local schools to handle local learning decisions, (g) ensure accountability measures for schools meeting standards, (h) reward high-performing schools, and (i) help low-achieving schools.

The vision of the 1990 Kentucky educational reform was grounded in six learning goals: (a) students will use basic communication and mathematics skills for real world scenarios, (b) students will develop academic core concepts from the required subject areas and apply the concepts to real life, (c) students will develop abilities to become self-sufficient, (d) students will develop attributes to function as responsible citizens in a variety of social structures, (e) students will develop critical thinking skills for real world problems, and (f) students will build on prior knowledge to acquire new learning (KDE, 2012c).

Under the new accountability system of the Kentucky Education Reform Act, KDE assessed all schools all schools to gauge the overall quality of that school. To determine the health of a school, the KDE developed the Kentucky Instructional Results Information System (KIRIS) as a standardized testing system to approximate growth toward learner goals. Eight years later, KIRIS was replaced by the Commonwealth Accountability Testing System (CATS). CATS combined content area assessments, norm-referenced assessments and additional results (e.g., graduation rates, attendance) into an Accountability Index. The Academic Index was one subscore of the Accountability Index and measured the criterion-referenced tests, known holistically as the Kentucky Core

Content Tests (KCCT). The KDE determined that all schools should have an Accountability Index score of 100 by the year 2014. Schools ranked as one of the following descriptors, in relationship to an Academic Index, under the broader Accountability Index: (a) meeting goal, (b) progressing, or (c) needing assistance.

By 2000, the goal of equitable educational opportunity for all students was realized: “Kentucky no longer has “rich” and “poor” districts,” as all schools in Kentucky are supported by roughly the same number of dollars (KDE, 2000b, p. 10). Financial equity among districts and high accountability for high standards are now the norms for Kentucky education (KDE, 2000b). However, Kentucky schools are still striving for educational excellence by incorporating measures from NCLB and the Race to the Top Grant.

Standards and Indicators for School Improvement

In a continuing effort to maintain a high level of achievement in learner outcomes for schools, the Kentucky Department of Education (2000b) developed the *Standards and Indicators for School Improvement* (SISI), representing a standards-based guide to the whole-school reform and improvement enacted in 1998 (703 Kentucky Administrative Regulation 5:120). The KDE then developed the Scholastic Audit to assess schools through a performance evaluation of the SISI. There were 88 indicators across the nine standards, with each of the standards constituting one element of whole school reform. The nine standards are divided into three sections: (a) Academic Performance, (b) Learning Environment, and (c) Efficiency. (A complete list of the SISI is located in Appendix A.) Throughout the current research, the author will refer to “Academic Performance” in the same manner as KDE, in so far as “Academic Performance” represents the first three standards of the SISI, and not as a common measure of school and district achievement. The current research focuses on the Academic Performance

standards, as listed in *School Level Performance Descriptors and Glossary for Kentucky's Standards and Indicators for School Improvement* (KDE, 2008):

Standard 1: The school develops and implements a curriculum that is rigorous, intentional, and aligned to state and local standards. (p. 3)

Standard 2: The school utilizes multiple evaluation and assessment strategies to continuously monitor and modify instruction to meet student needs and support proficient student work. (p. 10)

Standard 3: The school's instructional program actively engages all students by using effective, varied, and research-based practices to improve student academic performance standards. (p. 18)

The Scholastic Audit

Schools designated as “needing assistance,” as determined by the Academic Index, a subcategory of the overall Accountability Index, were required to conduct a Scholastic Audit, using the SISI standards and indicators as a guiding document for reform toward the goal of proficiency by 2014—and the ultimate goal of continuous improvement. Upon the performance of a Scholastic Audit, the evaluation team employed a spreadsheet that listed the nine SISI standards. The spreadsheet was divided into four columns, each with a set of descriptors to detail the varying levels of implementation in 4-point variable descriptions for each indicator (Kentucky Association of School Councils [KASC], 2012). KASC (2012, p. 5) termed the SISI as “the backbone of the scholastic audit and the scholastic review process.” During the 2000-2001 and 2002-2003 school years, 245 Scholastic Audits and Reviews were performed by the KDE across all levels of schools. Trained by the KDE, Scholastic Audit teams were composed of a parent, a teacher, a school administrator, a district administrator, a university faculty member, and a Highly Skilled

Educator (B. C. Ennis, 2007).

The schools determined by the KDE to need assistance under the CATS accountability system were audited through the Scholastic Audit process. While all schools could apply for the Scholastic Audit, and some high performing schools did, the majority of schools audited were mandated to conduct an audit (schools determined as “needing assistance” by the KDE). The audit not only used academic achievement as a measure of school health, but, also, learning climate, efficiency, leadership, and culture (Redding & Walberg, 2008). According to Redding and Walberg, one lesson gleaned from the reform movement was that it was not the students that lacked capacity (p. 72). Those schools labeled as “needing assistance” were then further sub-divided into three categories: the lowest one-third of those “needing assistance” was termed Level 3 schools and were made to conduct a Scholastic Audit; the middle one-third of those “needing assistance” was labeled Level 2 and were given a Scholastic Review; and, the highest one-third of those “needing assistance” were called Level 1 and were asked to conduct an internal self-review (B. C. Ennis, 2007).

The Problem Defined

Since the 1983 report, *A Nation at Risk*, by the National Commission on Excellence in Education, educational reform has been at the forefront of national, state, and local policy. Spearheading the state-level educational reform movement since 1990, Kentucky set the bar for the framework educational reform would assume moving forward. Through the *Kentucky Education Reform Act of 1990*, the KDE enacted sweeping measures for accountability in the commonwealth—noted by Steffy in 1993 as “the most comprehensive, innovative reform legislation ever passed by any state in recent history” (p. xiii). Then, perhaps more than ever, Kentucky was poised to make a huge dent in the

inequities of funding and insufficiencies of education that plagued the state. This mission was heralded as accomplished by the KDE in 2000, ten years following the landmark Kentucky Supreme Court decision that precipitated the development of KERA—as financial inequities among counties were corrected and high accountability based on exacting standards was enacted (KDE, 2000b). However, even with all the purported success for the KDE and the reform movement in Kentucky (see Miller & Moore, 2006), achievement still lagged behind the goal of “Proficient” for all schools by 2014 and inequalities still remained. Reaching proficiency and beyond on accountability measures represents one of the fundamental objectives for each school and district.

While Kentucky led the nation in the reform movement with the visionary measures outlined in KERA, the methodology regarding how to reach this goal of proficiency was left largely to individual school districts. Several years removed from the initial pedagogical efforts that marked the journey toward proficiency for school districts, a dearth of research still exists on what actually impacted academic performance, as it relates to elementary school science achievement. In the current section, the researcher will assess the aforementioned problem through four lenses: science accountability, school improvement tools, academic engagement, and the Scholastic Audit in science—each as pertaining to elementary science.

Science Accountability

Studies of science accountability are relatively rare compared to math and language arts; yet, science is one of the areas where United States school performance lags the most behind. In terms of recent national achievement on the science component of the 2011 National Assessment of Educational Progress (NAEP), “just 32% [of American students] were rated proficient or better, while a third were rated “basic” (indicating partial mastery

of the material) and 35% were below basic” (DeSilver, 2015, p. 4). From the varying extant literature on achievement gaps (Evans, 2005), students are observed to suffer a widening achievement gap for a variety of school, cultural, and socioeconomic-based factors while progressing through the traditional kindergarten to twelfth grade levels. Elementary science provides the scientific foundation of knowledge and application that is later expanded and assessed at the middle and high school level. No current study addresses the deficit of research for science accountability as it pertains to the Kentucky elementary classroom, when measured through SISI data from the Scholastic Audits.

School Improvement Tools

Although the SISI and Scholastic Audit have been recognized as perhaps the most effective school improvement tool in the nation (Mintrop, 2003; Mintrop & Trujillo, 2005), there are relatively few studies that have been conducted in this area of accountability. To date, the series of dissertations directed by Stephen Miller and Doug Smith (see B. C. Ennis, 2007; McKinney, 2007; Saravia, 2007) at the elementary level and Todd (2010) at the secondary level, as well as some overall work by Smith and Miller (2013) regarding measures of rurality, represent the only studies which examine the relationship between various combinations of the nine standards on school accountability outcomes. Even though the Scholastic Audit, in its original form and intent, is now defunct, the process of the Scholastic Audit and its basic framework under the SISI has now evolved with the Standards and Indicators Scholastic Review (SISR) work of Miller, Chon, Houchens, Smith, and Hunt (2014b). Thus, more research is needed utilizing the Scholastic Audit and the SISI to further the field.

Academic Performance

Of the extant studies on Academic Performance, McKinney (2007) examined

Standards 1 (Curriculum) and 3 (Instruction), but as mediating variables for Standard 7 (Leadership). Todd (2010) replicated this combination at the secondary level. Only one study (Smith & Miller, 2013) looked at the contribution of SISI Standards 1-3, collectively known as the Academic Performance standards, but their focus was on the effect of rurality on the entire Scholastic Audit process. Thus, more research is warranted to assess the possible correlations between the three Academic Performance standards of the SISI to school accountability.

Scholastic Audit in Science

Currently, no studies have been conducted on Kentucky accountability focused exclusively on elementary science, generally or specifically, using the school improvement tools of the SISI with the Scholastic Audit. In order to further push the education field toward growth and achievement in science, more research is needed to examine the confluence of science accountability, school improvement tools, academic engagement, and the Scholastic Audit impact on the Kentucky classroom—specifically, elementary science. While other empirical studies (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2007; Todd, 2010) have examined the SISI in terms of student achievement, school administrator impact, and professional development and leadership, no current study links the three Academic Performance standards of the SISI directly to elementary science accountability.

Purpose

This study brings together the issues described in **The Problem Defined** above. First, the researcher will examine the role of science accountability, as evidence of low international rankings for American science preparation and subsequent achievement, in preparing students to assume the mantle of Science, Technology, Engineering, and Math

(STEM) professions available, continues to trend in the first part of the 21st century. Second, with the implementation of school improvement tools such as the Scholastic Audit and the SISI, to guide the educational accountability in Kentucky, more studies are warranted to further assess the value of these aforementioned tools. Third, the research will assess the impact of the collective Academic Performance standards (Standards 1-3: Curriculum, Classroom Evaluation and Assessment, and Instruction, respectively) from the SISI. Fourth, with the dearth of science studies performed on Kentucky elementary schools, using the improvement tools of the Scholastic Audit and SISI, the current research will shed light on a deficit in the field of Kentucky accountability in science.

Thus, the purpose of this study is to examine the impact of the first three standards of the SISI document—Curriculum, Classroom Evaluation and Assessment, and Instruction—on elementary school science accountability scores for the Kentucky Core Content Test (KCCT). These three standards are collectively known as the Academic Performance standards of the SISI and seek to operationalize and approximate the direct impact of academic performance on student achievement. This study addresses the lack of current empirical data for the elementary classroom regarding accountability science scores in Kentucky. In the study, the researcher will analyze archival scores from the KCCT to gauge possible correlations between pedagogy in Kentucky elementary schools—as represented by the first three standards of the SISI, the Academic Performance standards—and science accountability scores. The study uses a correlational research design with a secondary database. The sample for this study was 181 audited elementary schools from the five rounds of Scholastic Audits (2001-2005) in Kentucky, from a total population of 669 Kentucky elementary schools (KDE, 2013). The researcher checked for repeat audits and found 17 instances and removed the “repeat incidents,” since the design assumes

independence. The removal of repeat incidents of audits is a unique feature of this research. The school is the unit of analysis for this study. The dependent variables are Kentucky CATS results, as measured specifically through elementary science achievement scores on the KCCT. The Academic Performance Standards 1, 2, and 3 of the SISI represent the target independent variables, Curriculum, Classroom Evaluation and Assessment, and Instruction, respectively. These three standards are a direct measure of the impact of the classroom academic performance. This research will also control for demographic data by incorporating a selection of school-level demographic independent variables. Following calculations of descriptive statistics and psychometric evaluation of the standards, correlational analysis utilizing multiple regression will be conducted.

Thus, this empirical study not only extends the literature on academic achievement and elementary science, but simultaneously analyzes the various relationships between accountability and elementary science in Kentucky, through the Scholastic Audit and SISI reform document. The central research question is, To what extent is Academic Performance (Standards 1-3 from Kentucky's *Standards and Indicators for School Improvement*) related to elementary school science accountability scores?

Research Questions

In this study, the researcher examines the relationship between school quality, as measured by Standards 1, 2, and 3 of the SISI, and elementary science achievement scores, as measured by the KCCT for 4th grade science—while controlling for demographic factors. As represented by Figure 1, the empirical research questions that constitute these relationships follow:

To what degree do(es):

1. Demographic Factors relate to Academic Performance (Standard 1, Standard 2,

Standard 3) and to KCCT Science?

2. Academic Performance (Standard 1, Standard 2, Standard 3) relate to KCCT Science?
3. Academic Performance (Standard 1, Standard 2, Standard 3) relate to KCCT Science, controlling for the Demographic Factors?

The SISI document provides a framework for this research, as observed in Figure 1.

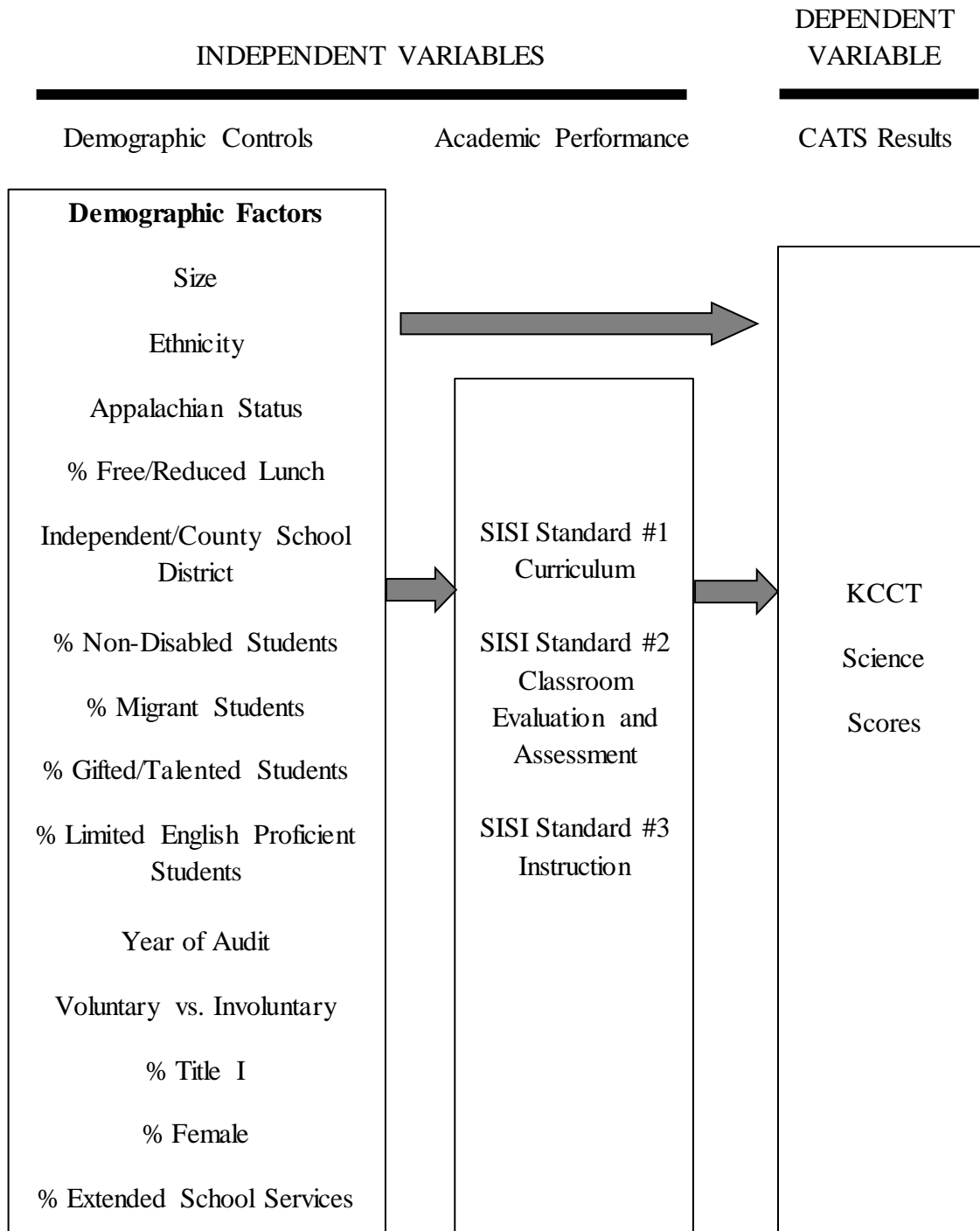


Figure 1. Hypothesized relationships among Demographic Factors, Academic Performance (Curriculum; Classroom Evaluation and Assessment; Instruction), and Science Accountability Outcomes.

Significance of the Study

In terms of significance, this study contributes to the areas of accountability research in several ways: science accountability, school improvement tools, academic performance, and Scholastic Audit research.

First, the study significantly extends the growing amount of empirical research on science accountability through the examination of school quality and student science outcomes. By examining academic factors and their subsequent contributions to school accountability, as measured by the elementary science component of the Kentucky Core Content Test (KCCT), this study may provide further insight regarding the impact of the classroom as it pertains to elementary science education—even for other states conducting similar reform measures to enhance student academic outcomes. Since Kentucky elementary school teachers are content generalists, as opposed to their content-specific secondary level counterparts, an examination of the impact of Academic Performance standards from the SISI on elementary science achievement could help determine whether exemplar instructional practices transcend a single, respective content area. Additionally, the current research could shed light on best practices specifically in science that if implemented would play a role in the reduction of gaps in earlier grades such as elementary school.

Second, this research further analyzes the school improvement tools of Kentucky's *Standards and Indicators for School Improvement (SISI)* and Scholastic Audit—regarded by some researchers as the premier school reform tools in the nation (Mintrop, 2003; Mintrop & Trujillo, 2005). The current study adds to the limited empirical base (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Smith & Miller, 2013; Todd, 2010) utilizing these two school improvement tools (Kentucky's SISI and Scholastic Audit).

Third, this study examines the role of academic performance on school accountability (the three Academic Performance standards from the SISI: Standard 1, Curriculum; Standard 2, Classroom Evaluation and Assessment; and, Standard 3, Instruction) along with demographic characteristics to elucidate any significant relationships. This research is the first to focus solely on the effects of just these three standards for Kentucky elementary schools.

Fourth, this research provides Kentucky educators and policy-makers correlations between school quality and KCCT science achievement scores. The study is the first and only analysis of selected standards from the SISI, as measured by the Scholastic Audit, for a separate academic discipline, in this case science, as extracted from the Accountability Index.

Limitations of the Study

The purpose of this study is: (a) to examine the effects of school quality, through the first three standards of the *Standards and Indicators for School Improvement (SISI)* document (representing Academic Performance: Curriculum, Classroom Evaluation and Assessment, and Instruction, respectively), on Kentucky Core Content Test (KCCT) elementary science accountability scores, and (b) to ascertain relationships, if any, among demographic independent variables such as size, ethnicity, Appalachian status, and student achievement. The quantitative data from this study was extracted from a secondary database at the Kentucky Department of Education (KDE) through the administration of the Scholastic Audits. The unique nature of Kentucky Scholastic Audits, in terms of location and variables, places several limitations on the current research. In addition, the extraction of data from a secondary database limits the current research in several ways.

First, while high-performing schools could elect to have a Scholastic Audit

conducted, the majority of Scholastic Audit scores in the sample are from low-performing schools who were mandated to conduct a Scholastic Audit. Thus, the inclusion of data from more high-performing schools could affect the measures of central tendency with the data currently listed.

Second, the current research only examines elementary science scores. Including content-based instruction and environmental factors, middle schools and high schools present a wide range of other factors outside of the elementary school context. Thus, generalizability to middle and high school levels would be subsequently limited.

Third, since the current research focuses only on elementary science scores, assessed during the fourth grade year, trying to reconcile the unit of analysis of the entire school—as Scholastic Audit teams measured the SISI standards for the entire faculty—versus the perception of only the fourth grade teachers could pose as another limitation.

Third, since the current study only investigates science scores, as opposed to the entire Academic Index (a multi-disciplinary accountability measure), generalizability to other subject areas could be limited.

Fourth, Scholastic Audit teams examine Academic Performance through the lens of multiple disciplines. A science-specific Scholastic Audit could provide different results but is beyond the scope of the current study, with data subsequently pulled from a secondary database, and, thus, poses as a limitation.

Fifth, the Accountability Indices, the Scholastic Audit data, and the SISI document all are contextualized and specific to Kentucky schools. While results could still be somewhat generalized for states with similar reform measures, state assessments, accountability procedures, and demographic data, generalizability would still be limited.

Sixth, since different schools were audited by different teams of individuals, human

error could have affected any inconsistencies across the spectrum of the hundreds of audits performed in the early 2000s in Kentucky.

Seventh, more factors than just the three main SISI independent variables and the demographic control independent variables could have affected the data: context of accountability measures at the time of the Scholastic Audit, evaluating school quality based on only three variables (Curriculum, Classroom Evaluation and Assessment, and Instruction), the interpretation of the SISI instrument, and deriving significant value from a school year of instruction with one to two days of science testing in the spring.

Eighth, since the inception of No Child Left Behind (NCLB) accountability measures, the majority of each school day in elementary school focuses on the teaching of reading and math, to the detriment of other subjects such as science, social studies, and the arts (Hinde, 2005, p. 105). The SISI Academic Performance standards specifically focus on classroom and school-level factors, regardless of subject area. Therefore, Scholastic Audit findings, which use the SISI as a grounding framework, refer to the entirety of all curriculum taught in all subject areas—not just science.

Summary

Since the seminal call for education reform and change in *A Nation at Risk* (1983), billions of dollars have been spent across the United States on raising standards for all vested parties in education, enacting high-stakes accountability for teachers and administrators, and reducing inequities among school districts for states. According to Steffy (1993), the Kentucky Education Reform Act of 1990 represented a radical shift in the educational paradigm of the early to mid-1990s: “KERA not only created a holistic framework for restructuring the public schools of this state, but it was initially funded with enough resources to enable successful implementation of the new programmatic

initiatives” (Steffy, 1993, p. xiii). The comprehensive reform package of KERA gave Kentucky a chance—to right the wrongs presented in a lawsuit by 66 Kentucky school districts. Ten years later, the KDE reported success through equitable support for school districts and a commitment to high standards and higher accountability (KDE, 2000b, p. 10).

Through KERA, the Commonwealth Accountability Testing System (CATS) was established as a second iteration of accountability for Kentucky, eight years after the controversial Kentucky Instructional Results Information System (KIRIS). CATS not only used assessments to arrive at an Accountability Index, but, also, employed additional results such as graduation rates and attendance. By 2014, the KDE set an Accountability Index goal of 100 for all schools. Through the Accountability Index, schools falling in the bottom third were labeled as “needing assistance.” The lowest group of the “needing assistance” category of schools was then made to conduct a Scholastic Audit (B. C. Ennis, 2007). Through the Scholastic Audit process, schools were evaluated over the course of several days based on the *Standards and Indicators for School Improvement (SISI)*. The SISI was a school evaluation instrument composed of nine standards across 88 indicators. The complete list of the SISI is given in Appendix A. During the two years studied in this research, 245 Scholastic Audits and Reviews were performed by Scholastic Audit teams (B. C. Ennis, 2007).

This study examines the theoretical model set forth in Figure 1 which approximates proposed relationships between Standard 1 (Curriculum), Standard 2 (Classroom Evaluation and Assessment), and Standard 3 (Instruction)—commonly known as the Academic Performance standards—simultaneously upon the dependent variable of student achievement, as measured by elementary school science scores on the Kentucky Core

Content Test (KCCT), while controlling for demographic factors. Other elementary level studies have also examined the impact of standards in the SISI, but with different sets of standards, such as McKinney (2007) with Leadership, Curriculum, and Instruction; Saravia (2007) with School Culture and Parental Involvement; and B. C. Ennis (2007) with Leadership and Professional Growth, Development, and Evaluation, in addition to Smith and Miller (2013) on the entire set of nine standards as related to rurality. However, rather than examine student achievement as a whole measure from the Academic Index, this is the first study to examine the first three standards of the SISI in relationship to elementary school science achievement.

This study represents quantitative research into a secondary database of the KDE to analyze potential correlations between the three Academic Performance standards of the SISI (Curriculum; Classroom Evaluation and Assessment; and Instruction) and elementary school science scores—while controlling for demographic variables. In the age of accountability, the central research question summarizes possible insights: To what extent is Academic Performance (Standards 1-3 from Kentucky's *Standards and Indicators for School Improvement*) related to elementary school science scores?

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

In 1989, the Kentucky Supreme Court made a landmark ruling in Kentucky education, declaring that “each child, every child, in this Commonwealth must be provided with an equal opportunity to have an adequate education” (Kentucky Department of Education [KDE], 2000b, p. 10). Developed in response to a lawsuit from 66 Kentucky school districts, the *Kentucky Education Reform Act of 1990* (KERA) has held schools accountable for achieving high standards in six goals and academic areas (Reedy, 2010). For the first time in Kentucky education, schools had a level-playing field in finance, opportunities for remediation with at-risk children, and clarity of learner goals with high standards of accountability (KDE, 2000b). In response to the accountability movement, researchers for the past 20 years have examined the response by educators to meet these enhanced demands on education.

The purpose of this study is to examine the impact of the Academic Performance standards (Standard 1, Curriculum; Standard 2, Classroom Evaluation and Assessment; and, Standard 3, Instruction) of the *Standards and Indicators for School Improvement* (SISI) document on science accountability scores from the Kentucky Core Content Test (KCCT) for Kentucky elementary schools. The study uses a correlational research design with data that are extracted from a secondary database, maintained by the Kentucky Department of Education (KDE). The sample for this study was 181 audited elementary

schools, out of 669 total Kentucky elementary schools, during five rounds of Scholastic Audits from 2001 to 2005 (KDE, 2013). In this study, the school serves as the unit of analysis. Kentucky Commonwealth Accountability Testing System (CATS) results represent the dependent variable and, specifically, the KCCT Science accountability scores for the elementary level (determined during the 4th grade). The aforementioned Academic Performance Standards 1-3—Curriculum, Classroom Evaluation and Assessment, and Instruction—serve as the target independent variables. In addition to the Academic Performance standards, several school-level demographic factors serve as additional independent control variables. After calculations of descriptive statistics and psychometric evaluation on the standards, correlational analysis with multiple regression are conducted.

This literature review is selective and examines empirical articles and related theoretical work, position papers, and policy analysis to approximate the edge of current research at the intersection of key terms such as *accountability, achievement, academic performance, curriculum, classroom assessment, instruction, and elementary science*. In addition, the prior dissertations analyzing the KDE Scholastic Audit database (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) and other key citations were examined for citations that did not surface during the electronic searches. The review of literature was conducted through the University of Louisville and Western Kentucky University libraries and utilized the EBSCOhost and ERIC databases.

The remainder of this Chapter II contains a literature review that is organized into the following sections: **Accountability, Academic Performance, Elementary Science, and Academic Performance, Accountability, and Science Outcomes**. This Chapter II concludes with a **Summary**.

Accountability

Beginning with the U.S. Department of Education's *Elementary and Secondary Education Act of 1965* (ESEA), continuing with the U.S. Department of Education implementation of the *No Child Left Behind Act of 2001* (NCLB), and, most recently, evolving toward a new iteration with the recent Congressional reauthorization of ESEA, educational accountability has been at the forefront of American social policy for the past 50 years (ASCD, 2015; U.S. Department of Education, 2015a; 2015b). Yet, in the most prominent document spearheading educational accountability and reform, *A Nation at Risk* (National Commission on Excellence in Education, 1982) outlined the need for reform and accountability with heightened Cold War implications that manifested a clear need for change (Kress, Zechmann, & Schmitt, 2011, p. 190). Moving from "minimum competency" toward "input-focused" accountability, *A Nation at Risk* (National Commission on Excellence in Education, 1982) propelled the American educational system toward setting high expectations, and the allocation of resources, to meet these goals (Kress et al., 2011, p. 190).

In the early 1970s, Levin (1974) broadly defined an accountability system as "one that maximizes the utility or satisfaction of the relevant constituencies subject to resource (and possibly other) constraints" (p. 375). Levin continued by stating that the level of accountability was reflected in how well the system of accountability achieved optimum results (p. 375). Levin further described accountability as a "closed loop" system whereby responses meet demands, activities reflect those demands, activities generate outcomes, and then outcomes drive further demands (p. 375).

Levin (1974) identified four taxonomic strands for accountability in the literature: (a) as performance reporting, (b) as a technical process, (c) as a political process, and (d) as an institutional process (pp. 363-364). While detailing the strengths and underlying

assumptions of each accountability construct, Levin proffered there was a possibility that the four concepts could form sequential stages in a quasi-accountability continuum: performance reporting unveils certain deficiencies in the system; technical modifications are embedded in the current system to remedy the deficiencies; the modifications prove unsuccessful and the political process is used to alter who is controlling the modifications; and, finally, the entire institution is considered dysfunctional if the problems persist (pp. 372-373). Levin concluded his taxonomy of accountability by stating the following: “as the diagnosis of the accountability problem shifts, so does the underlying concept” (p. 374).

In terms of educational accountability, Levin (1974) noted that performance reporting, “a periodic report of the attainments of schools and other educational units,” is the most probable common understanding of accountability in education (p. 364). Levin suggested that this focus on educational outcomes is an easily targeted and understood pathway for vested individuals to determine the proficiency of current educational measures and programs in schools (p. 364). However, within the concept of performance reporting, Levin found two underlying assumptions that hinder performance reporting as a comprehensive means to define educational accountability: (a) the assumption that all vested parties agree on the common objectives of education and, therefore, all the educational outcome data will be useful for the constituencies; and (b) that the information itself, regarding educational outcomes, will drive future improved educational outcomes (p. 365). In regards to the first assumption of universally accepted objectives for education, Levin stated “one of the salient issues that arises is accountability to whom and for what” (p. 365). Linn (2003) echoed this sentiment in outlining the two questions that need to be answered when addressing accountability: (a) What counts? and (b) Who is held

accountable? (p. 3). Much like the underlying assumptions noted by Levin (1974) for performance reporting, Linn (2003) continued by citing that a broad interpretation of the first question (“What counts?”) is required so that certain goals are not emphasized over other outcomes (p. 3). Linn proffered that the political systems that have outlined accountability systems in the past have defined accountability far too narrowly (p. 4).

By the early 1990s, accountability moved further toward the performance reporting concept of “consequential accountability” with a focus on the following three elements: (a) definitive standards that are communicated to all vested stakeholders; (b) testing through assessments aligned to these standards; and (c) performance-based consequences (Kress et al., 2011, pp. 185-186). Fuhrman (1999) cited seven differences between the new methods of consequential accountability and the accountability noted in previous literature on archetypal accountability: (a) district and school success based on student performance rather than compliance; (b) schools serving as the unit of analysis for improvement; (c) continuous improvement models focusing on school pedagogy and structures; (d) innovative methods to measuring classroom performance; (e) more types and levels of accreditation being enacted; (f) public dissemination of school test scores; and (g) further consequences linked to school performance levels (p. 1).

As previously observed by Levin (1974), a focus on performance reporting provided an easily understood pathway for vested educational parties to determine the effects of accountability measures. According to Levin, performance reporting is the cogent understanding of accountability because it focuses on the achievement of schools and other stakeholders toward educational objectives (p. 364). Under this assertion, Levin noted the assumption that the “performance reporting interpretation is that information on educational outcomes is necessary to enable constituents to appraise the proficiencies of

the schools” (p. 364). Fuhrman (1999) validated this original contention of Levin (1974) by stating that new accountability systems, such as consequential accountability, focused more on the performance of student achievement, as measured at the district and school level (p. 6).

According to Kress et al. (2011), the consequential accountability movement began in the 1990s as nearly forty states enacted intensive accountability legislation into public schools and as the federal government passed the *Improving America’s School Act of 1994* (U.S. Department of Education, 1994; Civic Impulse, 2015). The *Improving America’s School Act of 1994* (IASA) contained prototype language that later helped establish the NCLB legislation with “federally mandated use of state assessments to measure students’ progress toward state standards” (Kress et al., 2011, p. 193). In the following section, the current review of literature examines the standards movement and achievement and the manifestations at both the federal and state levels.

The Standards Movement and Achievement

During the early 1980s, with the release of *A Nation at Risk* (National Commission on Excellence in Education, 1982), and the subsequent printing of 70,000 copies during its first year of publication, the American educational landscape moved expediently toward political machinations that produced systematic educational reforms (Kress et al., 2011, p. 190). In less than a decade, states and districts began a very intentional process of building educational accountability systems to meet the demands entailed upon America by *A Nation at Risk* (National Commission on Excellence in Education, 1982; Kress et al., 2011, p. 190). State and district reforms included the following: (a) raising school budgets, (b) lowering student/teacher ratios, (c) raising graduation standards, (d) requiring more science classes, (e) extending the school day and school year calendars, and (f) raising educator

salaries (Kress et al., 2011, pp. 190-191).

However, during the late 1980s and the 1990s, the educational community purported a growing dissatisfaction with the “input-focused” reform of the past decade: “this seismic philosophical shift changed the focus from reliance on certain inputs to a more results-oriented approach, in which the focus was on achievement of improved student results, or what may be called outputs” (Kress et al., 2011, p. 191). Standards-based reform measures, spearheaded at the state level, were marked by the following indicators: (a) instructional standards; (b) the degree of success in student mastery of these standards; and (c) incentive and punitive measures based on success for student mastery of standards at the school and district level (p. 191).

Perhaps, as described by Fuhrman (1999), the transition to consequential accountability in education during the late 1980s and early 1990s can be further viewed through the lens of motivation. Fuhrman discussed the motivational aspects of accountability and stated “new accountability systems can be motivating, especially in the presence of certain enabling conditions, and that internal, shared beliefs about the importance of student performance are likely to be key enablers” (p. 8). Through two studies on accountability systems by Kelley, Milanowski, and Heneman, in North Carolina’s Charlotte-Mecklenburg School District and Kentucky (as cited in Fuhrman, 1999, p. 6), Fuhrman identified a strong correlation between establishing student performance goals and enhanced teacher performance in meeting these goals. In addition to setting student performance goals, Fuhrman observed that at both North Carolina’s Charlotte-Mecklenburg School District and Kentucky study sites, teachers reported positive incentives from the personal satisfaction of helping students achieve enhanced scholastic outcomes, professional satisfaction in being recognized as a high performing

school, and, to a lesser extent, professional recognition through a monetary bonus (pp. 7-8). In addition, Fuhrman articulated that teachers were also moved through the negative motivators of increasing pressure and stress to improve academic outcomes of student achievement and fear of having their respective school labeled in an adverse fashion (p. 8).

While some states used standards-based reform measures solely for “diagnostic or tracking purposes,” other states during the early 1990s began incorporating standards-based reform measures, and the subsequent performance reporting of educational outcomes, as a means of consequential accountability for rewarding and punishing schools based on the degree of success for achievement (Kress et al., 2011, p. 192). Consequential accountability was grounded in the notion that incentives would drive schools to build structures for ensuring student achievement at the highest level (p. 192). This type of accountability was a novel reform measure since it held, for the first time, schools accountable for the learning (p. 192). Hess (2005) stated “the allure of performance-based accountability is easy to understand. Standards represent a public commitment that schools will teach all children a discrete body of knowledge and skills to a specified level of mastery” (p. 53). By the mid-1990s, almost every state had passed legislation for standards-based education with assessment to measure the mastery of those standards by students (Kress et al., 2011, p. 191).

Through the enactment of the NCLB (U.S. Department of Education, 2001), consequential accountability became an even more prominent feature of the American educational outlook. With the passage of the NCLB, this notion of consequential accountability irrevocably framed education in America: “since 2001, considerations of school reform have been dominated by performance-based accountability. No Child Left Behind (NCLB) has, for better or worse, changed the way policy-makers and educators

talk about educational performance, and think about educational challenges” (Hess, 2005, p. 53). NCLB (U.S. Department of Education, 2001), improving upon the efforts of IASA (U.S. Department of Education, 1994) in identifying failing schools, bolstered provisions for consequential accountability in moving the standards-based reform of 39 states toward a national model of standards-based reform (Kress et al., 2011, p. 193). According to Kress et al. (2011), “at the time it was passed, the No Child Left Behind Act of 2001 represented the high point of the federal government’s involvement in efforts to assure and improve the quality of American education” (p. 193).

While mandating that every state move children to proficiency in math and reading by 2014, NCLB (U.S. Department of Education, 2001) allowed flexibility in determining how students of each respective state could best achieve proficiency (Kress et al., 2011, p. 193). Under the provisions of NCLB, each state was mandated to set rigorous achievement standards and to demonstrate Adequate Yearly Progress (AYP) for all students, regardless of subgroup identification (Linn, Baker, & Betebenner, 2002). NCLB outlined the following requirements for schools to meet AYP:

1. States must develop AYP statewide measurable objectives for improved achievement by all students and for specific groups: economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency.
2. The objectives must be set with the goal of having all students at the proficient level or above within 12 years (i.e., by the end of the 2013-2014 school year).
3. AYP must be based primarily on state assessments, but must also include one additional academic indicator.
4. The AYP objectives must be assessed at the school level. Schools that have

failed to meet their AYP objective for 2 consecutive years will be identified for improvement.

5. School AYP results must be reported separately for each group of students identified above so that it can be determined whether each student group met the AYP objective.
6. At least 95% of each group must participate in state assessments.
7. States may aggregate up to 3 years of data in making AYP determinations.

(Linn et al., 2002, p. 4)

In examining the impact of the NCLB after four years of implementation (U.S. Department of Education, 2001), Azzam, Perkins-Gough, and Thiers (2006) offered four conclusions: (a) teaching and learning are changing; (b) state achievement test scores are rising; (c) NCLB effects are holding steady; and (d) the greatest impact is seen in urban school districts (p. 94). First, in terms of the changing effects on teaching and learning, Azzam et al. noted that districts are using achievement score data to inform instructional decisions about specific deficits and hiring instructional coaches to enhance classroom instruction by building teacher leadership capacity (p. 94). Also, NCLB required teachers to obtain further qualifications in the teaching profession to maintain a highly qualified status (p. 94). Second, in examining NCLB impact on state achievement test scores, Azzam et al. stated that while student test scores were rising, other factors such as state re-classification of what constitutes “proficiency” and district policies and procedures could also account for the achievement score gains (p. 94). Third, in analyzing the consistency of the NCLB impact, the authors determined the steadiness of the current impact could be attributed, to a greater extent, to the federal and state regulations that have enabled school districts to accomplish AYP more easily (p. 94). Fourth, in regards to the school district

demographics where NCLB is having the most impact, Azzam et al. determined that urban school districts were the greatest beneficiary of NCLB provisions (p. 94).

The following sections give further background on the standards movement in Kentucky. Accountability in Kentucky education and the Kentucky Education Reform Act (KERA) are examined.

Accountability in Kentucky education. In April of 1990, Kentucky signed into law a nine-pound, 906-page document of educational reform, entitled the *Kentucky Education Reform Act of 1990* (KERA), that positioned Kentucky at the forefront of educational news (Pipho, 1990, p. 662). Through the landmark legislation of the KERA, Kentucky not only pushed itself to the forefront of the national education standards-based reform movement, but, also, legislated rigorous punitive and incentive ramifications for schools, establishing itself as a state imposing consequential accountability. Put another way, Kentucky moved from a system based on a logic of confidence (students are assumed to learn if they sit in classrooms where the teachers and schools are fully accredited) to a value-added model in which student growth is determined by measured progress from whatever level they begin toward high standards set by the state (Miller, 1992). While not the only state enacting consequential accountability measures, Kentucky was indeed one of the first. More detail on accountability in Kentucky under KERA is presented in the next section.

The Kentucky Education Reform Act. In the focused and seminal work on KERA, *The Kentucky Education Reform: Lessons for America*, Steffy (1993) called the KERA legislation “the most comprehensive, innovative reform legislation ever passed by any state in recent history” (p. xviii). Steffy continued by stating that “KERA not only created a holistic framework for restructuring the public schools of this state, but it was

initially funded with enough resources to enable successful implementation of the new programmatic initiatives” (p. xviii). The enthusiasm and energy that permeated the recesses of the state legislature were perhaps captured best by House Majority Leader, Greg Stumbo, moments before the passage of KERA:

Ladies and gentlemen of the House, after much debate and deliberation, you now have before you what I hope will become the bill that will truly change Kentucky’s educational opportunities for children in rural Kentucky and those all over this state.... I think that what we have the chance to do today is to see that that [students perpetually falling behind] never happens again. That no child in Kentucky will ever have to look back and say that he or she didn’t receive that full educational opportunity that he or she was entitled to by the constitution of this state. (Steffy, 1993, p. 2)

In the years preceding the passage of KERA by the state legislature, Kentucky education was marked with unremarkable descriptors: high dropout rates, low teacher salaries, per pupil funds well below the national average, and inequity in school finance among the poorest and wealthiest school districts (Steffy, 1993, p. 1). Due to the inequitable funding of student education in Kentucky, a class action equity suit was filed in Franklin Circuit Court in 1988 by the Council for Better Education (formed in 1985), led by former Kentucky Governor Bert Combs, and composed of sixty-six school districts, seven boards of education, and twenty-two public school students (Steffy, 1993, p. 2). While the initial ruling by Judge Ray Corns stated the Kentucky education funding system was “discriminatory” and “inefficient,” the subsequent verdict by the Kentucky Supreme Court, under appeal, found the entire Kentucky educational system unconstitutional (p. 2). Soon after the Kentucky Supreme Court verdict, the Kentucky legislature was entrusted

with the responsibility of reforming the Kentucky educational system by the end of the next session for the General Assembly, scheduled to convene in January of 1990 (p. 3).

The Kentucky General Assembly formed a Task Force comprised of three subcommittees: curriculum, finance, and governance (p. 3). Subsequently, what started as a class action equity suit in 1985 became a reality on April 11, 1990 as Kentucky Governor Wallace Wilkinson signed House Bill 940 into law (pp. 2-3).

Under the reformed structure of the new Kentucky educational system, KERA listed the following components for the “essential and minimal” requirements of an efficient system of education:

1. Its establishment, maintenance, and funding are the sole responsibility of the General Assembly.
2. It is free to all.
3. It is available to all Kentucky children.
4. It is substantially uniform throughout the state.
5. It provides equal educational opportunities to all Kentucky children.
6. It is monitored by the General Assembly to assure that there is no waste, no duplication, no mismanagement, and no political influence.
7. Schools are operated under the promise that an adequate education is a constitutional right.
8. Sufficient funding provides each child an adequate education.
9. An “adequate education” is defined as one that develops the following seven capacities:
 - a. communication skills necessary to function in a complex and changing civilization

- b. knowledge to make economic, social, and political choices
- c. understanding of government processes as they affect the community, state, and nation
- d. sufficient self-knowledge and knowledge of one's mental and physical wellness
- e. sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage
- f. sufficient preparation to choose and pursue one's life's work
- g. skills enabling students to compete favorably with students in other states. (Steffy, 1993, pp. 5-6)

In designing the curriculum initiatives for KERA, the curriculum subcommittee framed the work around thirteen principles about learning (Steffy, 1993, p. 6). Steffy cited three of these guiding principles as the most compelling determinants for curriculum initiatives: (a) all students can learn at high levels; (b) the current, known amount of educational research can ensure student success at schools; and (c) what students learn should be consistent across the entire state (pp. 6-7). In moving toward the type of Kentucky school that ensured the aforementioned guiding principles, KERA required five dramatic changes in structure to the existing system of education (p. 7). First, Kentucky education transitioned from "input" measures toward "outcome" measures when determining accountability structures (p. 7). Furthermore, while school districts would still be accountable, the school served as the primary unit of analysis when assessing student accountability (p. 7). Second, under the provisions of KERA that required improvement in student achievement over a specified period of time, rewards and sanctions would be implemented to maintain school focus toward these continuous improvement goals (pp. 7-

8). Third, KERA required that schools move from the norm-referenced tests (such as the California Test of Basic Skills) of the past toward a school-based, performance assessment system that aligned to state standards (p. 8). Fourth, in terms of governance, KERA determined that the “locus of school district decision making” should transfer from the state and district level to a system of school-based decision making councils and, consequently, empowering teachers (pp. 8-9). Fifth, accounting for the enormous amounts of new knowledge and training needed for teachers to meet the requirements of KERA, the last structural change significantly increased the funding to provide professional development (p. 9). Taking into account the synergy requisite to incorporate both the guiding principles and structural changes of KERA, Steffy stated “each component of the restructuring initiative is directly related to other components, forming a school restructuring web” (p. 10). The aforementioned statement by Steffy aligns with Systems Theory—the theoretical framework for the current research explicated later in this review of literature.

The reform structure of KERA was underscored by the following six learning goals for students: (a) able to use basic communication and mathematics skills for a variety of real-world situations; (b) able to develop and apply skills and abilities from the fields of mathematics, science, arts, humanities, practical living, and vocational training for a variety of real-world situations; (c) able to develop these skills to become self-sufficient learners; (d) able to develop these abilities to become responsible members of the community, in both family and work; (e) able to develop these skills to think and solve real-world problems; and (f) able to develop these abilities to integrate and build upon past knowledge and learnings (KDE, 2012c). Essentially, the Kentucky reform movement was underscored by the philosophical vision that every action and opportunity in the classroom

should revolve around what students should know and do, with the underlying assumption that “all students are capable of learning” (KDE, 2015b, p. 1).

As mentioned in Chapter I of the current study, the Kentucky Department of Education (KDE) developed the Kentucky Instructional Results Information System (KIRIS) as the testing accountability system to measure the degree of student achievement and success at the school level—in moving from input-focused to outcome-based accountability through KERA. KIRIS collected data on student outcomes in grades 4, 8, and 12 through a combination of on-demand constructed response questions, performance events, and portfolios (Stecher, Rahn, Ruby, Alt, Robyn, & Ward, 1997, p. 87). Stecher et al. (1997) reported that KIRIS was the first statewide performance-based accountability system in America (p. 87). However, with questions of the validity and quality of KIRIS, KDE transitioned to the Commonwealth Accountability Testing System (CATS) eight years later (Stecher et al., 1997). According to KDE, much like the original goals of KIRIS, CATS was designed to “encourage and enable educators in each public school to increase the academic achievement of their students” (KDE, 2006a, p. 1). Through CATS, schools were charged with meeting annual goals within the Kentucky Accountability Index. The Accountability Index was composed of three measures: (a) criterion-referenced tests of state standards; (b) a norm-referenced test; and (c) nonacademic indicators (KDE, 2006a). By the year 2014, all Kentucky schools were required to achieve the goal of proficiency, an Accountability Index score of 100 (KDE, 2006a).

Under the broader Accountability Index, the Academic Index was a subscore that measured the Novice, Apprentice, Proficient, and Distinguished (NAPD) performance levels (defined later in the **Definition of Terms** section of Chapter III) on the criterion-referenced tests of state standards. When the Academic Index was combined with the

norm-referenced tests and the nonacademic index factors (e.g., graduation rates, attendance), the overall Accountability Index was thus derived (KDE, 2006a, p. 49). According to the KDE, the variety of accountability measures used in the Accountability Index provided a “snapshot” of each school in order to “allocate resources and analyze instructional programs” (KDE, 2006a, p. 3). The criterion-referenced tests of state standards comprising the Academic Index was known as the Kentucky Core Content Tests (KCCT). The KCCT is defined and described in further detail in the section that follows.

KCCT test. The Kentucky Core Content Tests (KCCT) were constructed as a criterion-referenced test of Kentucky state standards and measured the Novice, Apprentice, Proficient, and Distinguished (NAPD) performance levels of students in the following areas: Reading, Math, Science, Social Studies, Arts and Humanities, and Practical Living/Vocational Studies (KDE, 2006a). According to KDE (2006a), Content Advisory Committees, comprised of eight to ten Kentucky teachers, advised in the construction of the multiple-choice items and open-response questions for the KCCT (p. 27).

According to the *2006 CATS Interpretive Guide* (KDE), the following division of questions were allotted for the KCCT. For the reading, mathematics, science, and social studies portions of the test, seven open-response and twenty-eight multiple choice questions were administered to students. In addition, one open-response and four multiple-choice questions were used as pre-test items and subsequently omitted from student scores and accountability calculations. For the arts & humanities and practical living/vocational studies tests, three open-response and twelve multiple-choice questions were administered to students. Additionally, one open-response and four multiple-choice questions were for pre-test items and excluded from student scores and accountability calculations. Taking into account the six total forms of the reading, mathematics, science, and social studies

tests, a total of thirty-six open-response and 144 multiple-choice items were used for accountability purposes. Additionally, when accounting for the twelve total forms of the arts & humanities and practical living/vocational studies tests, a total of twenty-four open-response and ninety-six multiple-choice items were administered for accountability purposes.

Regarding the two types of testing items on the KCCT test forms, the open-response questions and the multiple-choice questions each uniquely gauged student mastery of standards for Kentucky's *Core Content for Assessment 3.0* (KDE, 2006a). In terms of open-response items, these written, constructed responses by students were scored by hundreds of trained scorers. Taking nearly two months to score the tens of thousands of open-response items by students, KDE believed that student success on the open-response items was a direct reflection of students receiving high-quality instruction in the classroom (KDE, 2006a). Each open-response item was scored with a unique scoring rubric using a 0 to 4 scale. However, while higher scores help predict proficiency, the scores of 1, 2, 3, or 4 did not directly translate to the NAPD descriptors. According to KDE, a score of 0 corresponded to a student answer on an open-response item that was left blank or only restated the question. As the score moved toward a 4 on the scale, a student response cited more relevant information through correct and articulate language and thus reflected a higher level of understanding to the assessed standard on the item. In regards to the multiple-choice items, student responses on these items were scored either incorrect or correct on a *p*-value scale from 0 to 1, respectively. In addition, the multiple-choice items afforded KDE an opportunity to examine the coverage of content instruction in the specific domain and enhance reliability of the assessments.

When calculating student performance for a specific content area KCCT

assessment, the two components of each test, multiple-choice items and open-response questions, were weighted specifically to account for the emphasis KDE placed on each item type. For the KCCT assessments in reading, mathematics, science, social studies, arts & humanities, and practical living/vocational studies, sub-domain accountability was calculated to give educators specific information on the level of student mastery for groups of standards within the respective content area, at the school and state level. When calculating the sub-domain accountability for content areas of the KCCT, the multiple-choice items changed from the normal *p*-value statistic of 0 to 1 (incorrect or correct, respectively) to the open-response item raw-scale score of 0 to 4. To reflect this change in calculation, KDE devised the following formula to calculate the sub-domain total for any content area:

$$\text{Sub-Domain Total} = 1/3 (4 \times \text{MC Total}) + 2/3 (\text{OR Total})$$

where MC are the multiple-choice items; and OR are the open-response items (KDE, 2006a, p. 25). As previously mentioned, weighting the open-response items 2/3 of the entire sub-domain accountability reflected the belief of KDE regarding the importance of open-response items in determining student mastery of content (Riley, 2016).

During the summer of 2001, the Kentucky Board of Education (KBE) established new standards for Kentucky that were subsequently aligned to the assessments used for accountability in CATS (KDE, 2006a). According to KDE, while the cut points for determining student performance remained constant, the percentiles associated with student performance on the KCCT subsequently shifted to demonstrate student growth (KDE, 2006a). In determining the cut points for non-performance, medium Novice, and high Novice, scorers considered the entirety of student responses and the test items administered. Similarly, scorers implemented the same process for determining cut points

at the low, medium, and high Apprentice levels (KDE, 2006a).

For the current research, in examining the impact of the Academic Performance standards 1-3 (Curriculum, Classroom Evaluation/Assessment, and Instruction, respectively) of the *Standards and Indicators for School Improvement (SISI)* document for Kentucky on elementary science scores, the KCCT for elementary science was administered during the end of fourth grade (KDE, 2006a). As previously mentioned, the KCCT science assessment for a given student contained six open-response and twenty-four multiple choice items (Riley, 2016). Across the six different but parallel forms that a student could receive, the KCCT elementary science assessment included a total of thirty-six open-response items and 144 multiple-choice items for accountability calculations (KDE, 2006a).

With the enactment of Senate Bill 1 during the 2009 Kentucky General Assembly, Kentucky began the process of reforming its accountability assessments, moving from the KCCT to the Kentucky Performance Rating for Educational Progress (K-PREP) tests, to begin during the 2011-2012 school year (KDE, 2015a). In the following section, the K-PREP test is discussed in further detail.

K-PREP test. Enacted with Senate Bill 1 during the 2009 Kentucky General Assembly, the K-PREP tests reflected a shift from the content-focused criterion model of the KCCT tests. According to the Legislative Research Commission (2013), in an analysis of the KCCT and other forms of assessment, the KCCT did not accurately and comprehensively measure the entire spectrum of student achievement, with the test ceiling not extending to gauge the skills beyond proficient and distinguished performance descriptors (p. 36). The K-PREP assessment program was designed as a blended model of norm-referenced testing and criterion-referenced testing. Unlike its KCCT predecessor

that was comprised of only open-response and multiple-choice items, the K-PREP tests included multiple-choice, extended-response, and short answer items (KDE, 2015a). The K-PREP assessment program was designed to be administered for grades three through eight. The K-PREP captured a broader range of student performance than its KCCT predecessor, at both the low-performing and high-performing ends of the academic spectrum (Legislative Research Commission, 2013, p. 36). All components of the K-PREP were constructed by an outside vendor, NCS Pearson, with national norms for the norm-referenced test and customized Kentucky norms for the criterion-referenced testing (KDE, 2015a). The K-PREP assessment model was implemented during the 2011-2012 school year. For the current research, and its subsequent focus on the 2001-2005 testing cycle sample, only secondary data specific from the KCCT portion of the CATS accountability model were analyzed. However, K-PREP assessment information was included in the current review of literature to exemplify the Kentucky commitment toward continuous improvement in developing increasingly rigorous forms of accountability measures.

School Improvement in Kentucky

While much of the educational work in Kentucky during the 1990s was dedicated to bringing to scale the massive scope of the KERA vision, educational stakeholders soon determined that a shift from an awareness of input-based factors to enhancing outcome performance measures in actuality was crucial for moving forward in the KERA work (Todd, 2010). By 1998, in an effort to improve continuously the accountability model outlined in KERA, the Kentucky General Assembly reauthorized KERA with Kentucky Revised Statute 158.6455 to reflect a significant change toward a state focus on student learning (Todd, 2010; Bowles, Churchill, Effrat, & McDermott, 2002). Through the

provisions of modified accountability in KRS 158:6455 (Section 3), consequential accountability was realized through punitive measures for schools that failed to meet student performance standards (Koger & Thacker, 2004, p. 1). Additionally, as a result of KRS 158:6455 (Section 4), the Scholastic Audit process was born with prescriptive guidelines for audit teams to help failing schools meet improvement goals (Koger & Thacker, 2004, p. 1). The Kentucky Department of Education (KDE, 2000b) developed the *Standards and Indicators for School Improvement* (SISI) to guide efforts toward whole-school improvement, with the Scholastic Audits as a way of measuring the SISI (703 Kentucky Administrative Regulations 5:120). In the following sections, the SISI are further explicated in its nature, development, application, and how it informed the Scholastic Audit process.

Standards and Indicators for School Improvement. Per the legislation of 703 Kentucky Administrative Regulation 5:120 and its focus on enhancing the whole school reform model, the *Standards and Indicators for School Improvement* document served to guide the school improvement for struggling schools during a Scholastic Audit. The SISI included nine standards with 88 indicators that directed the Scholastic Audit in its areas of focus. Each of the nine standards reflected an individual component of best practices for whole school reform, divided into three domains: (a) Academic Performance, (b) Learning Environment, and (c) Efficiency. The Academic Performance standards encapsulated the school improvement efforts that focused on the individual classroom and included the following: Standard 1, Curriculum; Standard 2, Classroom Evaluation/Assessment; and Standard 3, Instruction. For the entirety of the current study, the researcher uses the term *Academic Performance* to reference the academic activities occurring at the classroom level; therefore, Academic Performance in this study should not be confused with the final

product of school and district achievement measures. The Learning Environment standards focused on the school culture and learning climate for the respective school and included the following standards: Standard 4, School Culture; Standard 5, Student, Family, and Community Support; and Standard 6, Professional Growth, Development, and Evaluation with two substandards: Professional Development, and Professional Growth and Evaluation. The Efficiency standards outlined improvement efforts at the administrative level and included the following: Standard 7, Leadership; Standard 8, Organizational Structure and Resources; and Standard 9, Comprehensive and Effective Planning (a complete list of the SISI is located in Appendix A).

The current research examines the impact of the first three SISI standards, Academic Performance, on elementary science KCCT scores from CATS. As listed in *School Level Performance Descriptors and Glossary for Kentucky's Standards and Indicators for School Improvement* (KDE, 2004b), the Academic Performance standards are the following:

Standard 1: The school develops and implements a curriculum that is rigorous, intentional, and aligned to state and local standards. (p. 3)

Standard 2: The school utilizes multiple evaluation and assessment strategies to continuously monitor and modify instruction to meet student needs and support proficient student work. (p. 10)

Standard 3: The school's instructional program actively engages all students by using effective, varied, and research-based practices to improve student academic performance standards. (p. 18)

Development of the SISI. In response to legislation in KRS 158.6455 (Section 4), which mandated the development of procedures for conducting Scholastic Audits, the

Kentucky Department of Education (KDE) used the *Standards and Indicators for School Improvement* (SISI) as a means for identifying focus areas during the school improvement process of the Scholastic Audits (Koger & Thacker, 2004, p. 1). Todd (2010) stated that the initial evaluation of the SISI was performed by the National Study of School Evaluation (NSSE), an educational research organization in Chicago. Anchored around nine standards and 88 indicators, the SISI guided Scholastic Audit data teams in facilitating focused school reform.

The SISI document included a four-column performance evaluation rubric by which each indicator could be referenced on a 1-4 scale, with behavioral descriptors of the Scholastic Audit for each “ratings of performance” measurement in the SISI. The 1-4 variable scale referenced the following “ratings of performance”:

Category 1—Little or no development and implementation

Category 2—Limited development or partial implementation

Category 3—Fully functioning and operational level of development and implementation

Category 4—Exemplary level of development and implementation. (KDE, 2004b, p. 3)

While the KDE conducted no validity studies specifically on the SISI, the current research aligns with other studies (see dissertations by B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) plus further analyses by Smith and Miller (2013) that have been conducted using the SISI as the primary instrument for school improvement. Each of these studies included validation of selected standards utilizing factor analysis and external criterion validity coefficients.

Regarding reliability, Koger and Thacker (2004) performed a validity assessment

on the Scholastic Audit process, with the SISI being the central document. In addition, Scholastic Audit teams received training on effective practices for using the standardized instrument of the SISI (Koger & Thacker, 2004, p. 3, 80). However, Koger and Thacker (2004) and the Legislative Research Commission (2005) cited several limitations to inter-rater reliability which will be discussed in a following section in the current review of literature on Limitations and Potential Problems.

The SISI document of nine standards and 88 indicators guided Scholastic Audit teams based on efforts by KDE to improve low-performing schools. In addition, Lyons and Barnett (2011) stated that KDE used the Scholastic Audit process, and the subsequent SISI document, for not only low-performing schools, but, also, for schools meeting their goals hoping to glean more insight into the school improvement process. Therefore, in an effort to focus on the specific distinguishing characteristics between the low and high-performing schools, KDE (2003) authored an analysis of these distinguishing characteristics. Originally termed *leverage points*, KDE (2003) later named these characteristics *variance points* since they represented the variance in Scholastic Audit results between the two groups of schools (Lyons & Barnett, 2011, p. 1). According to Lyons and Barnett, the *variance points* concept was of particular interest to many Kentucky stakeholders since it sought to clarify the distinctive elements between low and high-performing schools (p. 1). KDE (2003) reported 27 variance points during its first round of comparisons between low and high-performing schools. These 27 variance points were subsequently used as suggestions for school improvement best practices (Lyons & Barnett, 2011, p. 4). During subsequent rounds of the Scholastic Audits, other variance points, ranging in number from 11 to 51, were determined from the aforementioned comparisons of schools (p. 4).

Unfortunately, the process by which KDE determined these leverage and later variance points was by “eyeball” rather than performing chi-square analysis (personal communication, Douglas Smith, December 4, 2015), a factor perhaps contributing to the inconsistent lists of which indicators rose to the level of distinguishing between high and low school outcomes from one set of audits to the next. The validation studies conducted under the oversight of Douglas Smith (dissertations cited above, plus Smith & Miller, 2013) focused on the standards and their respective sets of indicators (as the SISI was designed to be interpreted) rather than examining the separate effect of each of the 88 indicators irrespective of their placement in one of the standards. Thus, validating the standards as separate unitary constructs puts the emphasis of content back on the effects of the nine over-arching standards. This corrects the focus on separate indicators that resulted from KDE’s attention to indicator-level content with the variance and leverage point analyses.

Since the educational reform movement of the 1990s, many facets of educational accountability in Kentucky have been modified and revised. However, the SISI remained a constant force in guiding school improvement since its inception in 1998 (Todd, 2010). Through the meticulous application of the SISI, Scholastic Audit teams possessed a valid, reliable instrument to inform school improvement within the educational reform process in Kentucky (see B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010).

Application of the SISI. As mentioned in the previous sections, the *Standards and Indicators for School Improvement* (SISI) represented a landmark reform document for guiding whole school improvement in Kentucky. The language of the SISI standards embodied the KERA belief that all children can learn at high levels and was specifically geared toward invoking an attitude of high expectations for school staff and administrative

leadership: “effective...research-based, results driven...high performance expectations...[and] high student and staff performance” (KDE, 2008; R. Hurley, personal communication, October 21, 2015).

In addition to guiding the Scholastic Audit procedures for low-performing schools, the SISI served the main function of providing an overall framework for whole school improvement and, in conjunction with the *School Level Performance Descriptors for Kentucky’s Standards and Indicators for School Improvement* (2004b), allowed schools, regardless of current performance, to capture the emphasis on these school reform efforts in the comprehensive school improvement plans (CSIP).

In accordance with the SISI and the *School Level Performance Descriptors for Kentucky’s Standards and Indicators for School Improvement* (2004b), schools were afforded the opportunity to assess the current state of school effectiveness in the three domains of the SISI document—Academic Performance, Learning Environment, and Efficiency. In essence, schools could use both documents to evaluate, plan, and implement the elements of school improvement most needed and align these structures and practices in the CSIP. Additionally, the two documents served as an effective and efficient protocol for schools to apply and implement during the beginning, middle, or end of the school year, reflecting the most current and appropriate data, and therefore informing the most powerful and cogent school improvement measures (McKinney, 2007).

Within the nine standards of the SISI document, the current research focuses on the first three, the Academic Performance standards: Standard 1, Curriculum; Standard 2, Classroom Evaluation/Assessment; and Standard 3, Instruction. These three standards are reviewed later in the current review of literature.

Scholastic Audits. Upon passage of KRS 158.6455 by the Kentucky General

Assembly in 1998, the Kentucky Board of Education (KBE) was tasked, in Section 3 and 4 of the legislation, with developing consequences for not meeting accountability standards and for developing procedures to help failing schools meet these standards, respectively (Koger & Thacker, 2004, p. 1). From these measures, the Kentucky Department of Education (KDE) composed the *Standards and Indicators for School Improvement (SISI)* and embedded this document into the subsequent school improvement protocol of the Scholastic Audit, to aide schools in meeting the accountability measures of CATS.

Under the Kentucky accountability system, a Scholastic Audit was conducted for schools not meeting achievement goals set forth by KERA. The Scholastic Audit employed the SISI as the primary tool for evaluating schools (KDE, 2008). The evaluation consists of 88 indicators across nine standards. Each indicator has a set of four specific descriptions (behavioral anchors) to assess the level of each school's performance during the week-long Scholastic Audit.

In order to classify schools most in need of assistance, KDE developed a taxonomy for identifying schools in relationship to their performance through CATS. With all schools required to meet the proficiency goal of 100 by 2014, KDE constructed a timeline in 2000 that set each school on a path toward proficiency through seven two-year cycles (bienniums), starting from the initial and respective baseline Accountability Index. One main measure of the Accountability Index was the Academic Index, a subset of the overall Accountability Index that included school performance on the criterion-referenced Kentucky Core Content Tests (KCCT) and the norm-referenced Comprehensive Test of Basic Skills (CTBS). KDE assigned schools the following identification based on their Academic Index: (a) meeting goal, (b) progressing, or (c) needing assistance. For those schools classified under the lowest descriptor, "needing assistance," KDE further grouped

these low-performing assistance schools in the following three groups: (a) Level 1, the upper one-third of schools below the assistance line; (b) Level 2, the middle one-third of schools below the assistance line; and (c) Level 3, the lowest one-third of schools below the assistance line (KDE, 2006a, pp. 16-17).

In terms of providing assistance for these three levels of schools needing assistance, KDE developed the following guidelines for meeting the needs of each grouping. First, regarding Level 1 Assistance schools, KDE prescribed a self-review using the SISI as the guiding document. Level 2 Assistance schools were required to conduct a scholastic review, which used a modified version of the Scholastic Audit procedures with only two members of a regional service team and two to four members from the respective school district. Level 3 Assistance schools, the lowest performing schools in the state, were fully audited by a KDE Scholastic Audit team that included the following composition, all from outside the audited school district: a parent, a teacher, an administrator, a district administrator, a university faculty representative, and a Highly Skilled Educator (HSE). The HSE was an audit team representative who was trained in the Scholastic Audit procedures, per the training protocol of KDE, and led the school reform efforts with the guidance of the SISI (Legislative Research Commission, 2006).

Koger and Thacker (2004) reported that 131 Scholastic Audit and Review procedures were conducted during the initial 2000-2001 accountability cycle. Furthermore, an additional 114 Scholastic Audits or Reviews were performed during the 2002-2003 school year. In both cycles of audits and reviews, a small sample of successful schools performed Scholastic Reviews for insight into what distinguished them from lower-performing counterparts. These distinguishing characteristics later informed the concept of *variance points* (Lyons & Barnett, 2011, p. 1).

Scholastic Audit and Review procedures focused on using the SISI document, and its three guiding domains of Academic Performance, Learning Environment, and Efficiency, to effect change in low-performing schools. When performing a Scholastic Audit, the trained Scholastic Audit teams collected current school evidence from the following sources: (a) leadership and culture surveys; (b) school portfolio artifacts; (c) team experiences; (d) interviews; and (e) observations (KDE, 2003). Examining the school evidence, Scholastic Audit teams assigned performance level variable descriptors, with the following points in consideration: (a) determining how thoroughly the evidence aligned with the performance indicators; (b) assigning the indicator scores holistically and resisting averaging; and (c) assigning scores for each respective performance indicator is not necessarily based on evidence available from all criteria statements (KDE, 2005b). In addition, Scholastic Audit and Review representatives employed a team approach when conducting the auditing functions with member composition defined in statute, training outlined in regulation, roles and responsibilities described in both statute and regulation, and a variable scoring rubric approved by the KBE (KDE, 2005b).

For the current research, only audited elementary schools are used as the sample ($N = 181$) from the five rounds of audits conducted between 2001 and 2005. In addition, the 4-point variable scores for each indicator have been modified to scores of 0, 2, 5, and 7 to represent the more difficult (greater stretch) school improvement transition from the “limited development or partial implementation” of Category 2 to the “fully functioning and operational level of development and implementation” of Category 3 (see McKinney, 2007; Saravia, 2008).

With the SISI as the touchstone document, Scholastic Audits offered an exacting, rigorous methodology for Kentucky educators to make school improvement gains in the

lowest performing schools, firmly grounded in the KERA framework for educational reform.

Limitations and Potential Problems

The *Standards and Indicators for School Improvement* (SISI) document was assessed for validity by Koger and Thacker (2004). However, the document itself did not serve as the primary basis for validation as much as how it served the function of driving the Scholastic Audit and Review process (Todd, 2010). The completed set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010), conducted under the counsel of Drs. Miller and Smith, offered substantial evidence for construct and external criterion validity of the application of the SISI document during Scholastic Audit procedures, with the parallel use of Cronbach's (1951) alpha for internal scale reliability. Additional studies have further analyzed the validity of the SISI instrument and confirmed it as unified and efficacious (McKinney, Miller, & Smith, 2007; Smith, B. C. Ennis, Saravia, Miller, & Wagner, 2008; Smith, Harvey, Hammock, & Miller, 2009; Todd, 2010). However, additional investigations of the SISI could serve as further validation of the document. The current research moves the field work forward through the use of the first three standards of the SISI, the Academic Performance standards as related to a different accountability measure, specifically, using only the elementary science scores from the criterion-referenced Kentucky Core Content Tests (KCCT) rather than the entire Academic Index.

When examining the reliability of the SISI document and the Scholastic Audit process, two potential problems should be noted. First, some confusion exists surrounding whether cross team and rater reliability were conducted on the Scholastic Audit procedures. Todd (2010) cited a personal communication that accounted for cross team

and rater reliability during the Scholastic Audits and reviews. However, Koger and Thacker (2004) stated that inter-rater reliability was never assessed on the SISI document and Scholastic Audit process. Lyons and Barnett (2011) also noted the absence of inter-rater comparison among different audit teams to bolster reliability (p. 11). Second, the Legislative Research Commission (2005) cited evidence from a 2002 HumRRO study that found that Kentucky schools, under the CATS accountability system, were accurately classified as meeting goals, progressing, or in need of assistance only 77% of the time (p. 44). Thus, Scholastic Audits and procedures, while otherwise reliable, could have been conducted on incorrectly identified schools. To ensure reliability, Koger and Thacker (2004) stated that audit teams received training on using the SISI during Scholastic Audits and performed the audits with a standardized instrument. However, the composition of audit teams and the Scholastic Audit as well as the KDE audit training changed annually (p. 80). Saravia (2008) noted however that validity was more important than reliability. Thus, in terms of grounding the research instrument and process, the SISI document can be considered a valid instrument for use during the Scholastic Audit process despite the inconsistencies noted here regarding reliability.

Academic Performance

Within the nine standards of the *Standards and Indicators for School Improvement* (SISI) document, the first three standards constituted the Academic Performance domain: Standard 1, Curriculum; Standard 2, Classroom Evaluation and Assessment; and Standard 3, Instruction. These three standards of Academic Performance encapsulated the school reform efforts that occurred within the classroom. The SISI assisted school leadership teams and educators by providing 88 indicators across the nine standards to guide improvement efforts. According to the Kentucky Department of Education ([KDE],

2005b), the standards in the SISI served as building blocks that could be purposefully arranged, re-arranged, and emphasized to impact change. With substantial research to support a laser-like focus on improving the seven indicators of Standard 1 (Curriculum), the eight indicators of Standard 2 (Classroom Evaluation/Assessment), and the eight indicators of Standard 3 (Instruction), low-performing schools could achieve considerable gains toward proficiency. While an extensive review of the current research on curriculum, classroom evaluation/assessment, and instruction is beyond the scope of this review of literature, the following sections provide an overview on these three classroom measures, as they align to the respective indicators for each SISI standard.

Curriculum

The *Standards and Indicators for School Improvement* (SISI) outlined nine standards that provided a framework for whole school reform in Kentucky. The first SISI standard, Curriculum, was defined by the Kentucky Department of Education (KDE) as “the school develops and implements a curriculum that is rigorous, intentional, and aligned to state and local standards” (KDE, 2004b, p. 3). Standard 1, Curriculum, included indicators to assess the current practices in the classroom that would subsequently impact academic performance for students. The following seven indicators constituted SISI Standard 1, Curriculum:

- 1.1a – There is evidence that the curriculum is aligned with the *Academic Expectations, Core Content for Assessment, Transformations, and the Program of Studies*.
- 1.1b – The district initiates and facilitates discussions among schools regarding curriculum standards to ensure they are clearly articulated across all levels (P-12).
- 1.1c – The district initiates and facilitates the discussions between schools in the

district in order to eliminate unnecessary overlaps and close gaps.

1.1d – There is evidence of vertical communication with an intentional focus on key curriculum transition points within grade configurations (e.g., from primary to middle and middle to high).

1.1e – The school curriculum provides specific links to continuing education, life and career options.

1.1f – There is in place a systematic process for monitoring, evaluation and reviewing the curriculum.

1.1g – The curriculum provides access to a common academic core for all students. (KDE, 2004b, pp. 3-9)

According to Lockwood (1994), Miller discussed the importance of learning as central to schools, occurring within the “iron triangle”—unchanging, inviolate—of the interaction of teachers with students as they engage the curriculum. With curriculum as one of the cornerstones of the “iron triangle” for education, the inclusion of curriculum by KDE as the first standard in the SISI reflects its grave importance in affecting school change. Farenga, Joyce, and Ness described curriculum as the “manner in which content is defined, arranged, and emphasized” including “structure, organization, and delivery of content” (as cited in Bybee, 2002, p. 52). Also, as referenced by the “iron triangle” of Lockwood (1994), Farenga et al. characterized curriculum as the conduit by which teachers and students interact (as cited in Bybee, 2002).

During a gathering of representatives from 21 high-performing Kentucky schools in 1999, including teachers, parents, principals, and superintendents from the respective districts, Clements reported the Prichard Committee for Academic Excellence and the Partnership for Kentucky Schools, who conducted the meeting, asked stakeholders their

beliefs on why each of these schools achieved at such a high level (as cited in Petrosko & Pankratz, 2000). After achieving “reward” status in all three accountability cycles of the Kentucky Instructional Results Information System (KIRIS) administration (1994, 1996, and 1998), Clements stated that the 21 school representatives of this meeting constituted more than half of the only 38 schools, out of approximately 1400 schools, that garnered this distinction: with 35 out of the 38 schools being elementary schools (as cited in Petrosko & Pankratz, 2000). According to Clements, after several hours of this meeting, the Prichard Committee for Academic Excellence and the Partnership for Kentucky Schools both agreed on one common strand that set these schools apart from their peers: “each school had made as its priority a well-articulated curriculum aligned with the state’s evolving assessment, and the adults in those schools focused on ensuring that every child had an adequate opportunity to learn the curriculum” (as cited in Petrosko & Pankratz, 2000, p. 98). The success of these high-performing schools with the indicators of the SISI document could be directly connected with Indicator 1.1a on Standard 1, Curriculum.

According to Petrosko, Lindle, and Pankratz (2000), KDE supported curriculum alignment and development efforts through the following four documents: the list of 57 Academic Expectations; the *Core Content for Assessment 1.0*; the *Transformations* document; and the *Program of Studies*. In terms of the 57 Academic Expectations authored by the Kentucky Department of Education, these expectations were used to guide curriculum development and were developed from the six learning goals of KERA, as referenced previously in the current review of literature. The 57 Academic Expectations further expounded upon four of the six learning goals of KERA, goals one, two, five, and six. Goals three and four from those learning goals were not used in the 57 Academic Expectations due to the difficulty of operationalizing student self-sufficiency and

becoming a responsible group member, respectively (Petrosko et al., 2000).

KDE authored *Transformations: Kentucky's Curriculum Framework*, a two volume collection that served as the curriculum framework for Kentucky educators, providing direction in curriculum development (Petrosko, 2000; Petrosko et al., 2000). Lindle (2001) noted that *Transformations: Kentucky's Curriculum Framework* was originally intended as a desk reference for teacher unit and lesson design. In addition, the Appalachian Educational Laboratory asserted that the massive document was rarely used due to its non-user-friendly nature (as cited in Lindle, 2001).

The *Core Content for Assessment* (KDE, 1996) document moved through several iterations, reflecting the KDE modifications and updates to the document since its inception in the mid-1990s. Regardless of the version, the *Core Content for Assessment* document provided detailed curriculum guidance for districts and schools and afforded a minimum framework of skills and knowledge that should be included for curriculum at the district and school level (Petrosko, 2000; Petrosko et al., 2000).

Another curriculum development assistance document, *The Program of Studies* outlined the minimum content standards that each student should master before graduating high school (KDE, 2006b). In addition to the requirements for high school graduation, the *Program of Studies* also vertically detailed the content standards for each level, including primary, intermediate, and middle grade levels (KDE, 2006b).

Building upon the common theme of curriculum alignment in high-performing schools during the aforementioned 1999 meeting, Clements also noted the intentional conversations of these schools regarding curriculum alignment across grade levels, a timeline of student mastery on performance standards, and how best to meet the needs of individual students (as cited in Petrosko & Pankratz, 2000). According to Clements, the

dedication of these 21 schools to take the first step toward school reform, through the alignment of the curriculum and the subsequent individualization of this curriculum, represented perhaps the apex of policy and practice, as intended by the Kentucky legislature when passing the KERA reforms nearly ten years prior (as cited in Petrosko & Pankratz, 2000).

Transitioning from inputs to outcomes as an educational paradigm has been at the forefront of American education since the late 1980s, early 1990s. Educational stakeholders focused less on the inputs of increased funding, the appropriate curriculum, and the right schedule and more on the intentional activities that produced results and outcomes, specifically, enhanced student achievement (DuFour, DuFour, Eaker, & Many, 2006; Fuhrman, 1999; Kress et al., 2011). For Standard 1, Curriculum, of the SISI, Indicators 1.1b, 1.1c, and 1.1d stressed the importance of districts initiating and facilitating inter- and intra-school level discussions regarding the clear articulation of P-12 curriculum standards, conversations to eradicate overlaps and gaps within the curriculum, and vertical conversations on curriculum transition points among grade level configurations, respectively (KDE, 2004b, pp. 4-6).

Since the early 2000s, DuFour et al. (2006) noted the inclusion of a Professional Learning Community (PLC) as representing one avenue for accomplishing these horizontal and vertical professional conversations at the district and school level. With no two opinions on educational impact alike, DuFour et al. stressed the value of these professional conversations inherent to the PLC process: “it is through the collective examination of results—tangible evidence of student learning—that teachers’ dialogue moves from sharing opinions to building shared knowledge” (p. 147).

In developing opportunities for curricular conversations between students and

teachers, educational curriculum developers often emphasized the concept of relevance, in an attempt to connect content to the varying avenues of life, college, and career for students. Embedded within supporting evidence documents for Indicator 1.1e for Standard 1, Curriculum, schools aligned the Comprehensive School Improvement Plan (CSIP), units of study, lesson plans, and field trips to reflect real-world opportunities for P-12 students (KDE, 2004b, p. 7). College and career readiness became a focus of state accountability in 2011, following the passage of Senate Bill 1 by Kentucky legislators (KDE, 2012a). Senate Bill 1 outlined two initiatives pertaining to college and career readiness: (a) to reduce by 50% the 2010 college remediation rates for high school graduates attending institutions of higher education and (b) to increase the completion rates for these college remediation courses by three percent annually, both goals respectively having 2014 benchmarks for completion (KDE, 2012a).

In addition to the development and alignment of curricular documents for districts and schools, KDE also outlined in indicator 1.1f for Standard 1, Curriculum, that systems and structures should be established for continuous monitoring of the curriculum (KDE, 2004b, p. 8). The bulk of the curriculum evaluation efforts in the late 1990s and early 2000s were conducted through the school Site-Based Decision Making Council (SBDM) and included in the respective school CSIP. As the 2000s progressed, the value placed on student achievement data and subsequent analysis of this data, two measures listed by KDE (2004b) as evidence for this indicator, multiplied exponentially. Yet, even with the proliferation of data analysis protocols, data retreats, and data teams, Reeves and other educational researchers questioned the value of what was being assessed: were students over-tested, yet under assessed? (as cited in White, 2011, p. 10). White (2011) continued by stating that analyses of student data are used more for ranking and comparisons than

improving actual student performance toward proficiency (pp. 10-11).

With its emphasis on “access to a common academic core for all students,” KDE (2004b, p. 9) authored Indicator 1.1g with the same cogent language that reflected the gravity of the original KERA phrasing that “all students can learn at a high level” (KDE, 1990). In addition, the concept of educational equity of a common academic core curriculum for all students was also addressed by KDE through the language in the performance rating 3 (“fully functioning and operational level of development and implementation”) on the SISI: “challenging curriculum...available to all students” and “elicits higher order thinking and problem-solving skills from all students” (KDE, 2004b, p. 9). In essence, Standard 1, Curriculum, and the subsequent seven indicators, bolstered this early 1990s KDE mentality, perhaps best addressed by the aforementioned former House Majority Leader Greg Stumbo, “that no child in Kentucky will ever have to look back and say that he or she didn’t receive that full educational opportunity that he or she was entitled to by the constitution of this state” (Steffy, 1993, p. 2).

Classroom Evaluation/Assessment

Classroom Evaluation/Assessment constituted the second SISI standard and contributed to the domain of Academic Performance. KDE (2004b) defined the second SISI standard, Classroom Evaluation/Assessment: “the school utilizes multiple evaluation and assessment strategies to continuously monitor and modify instruction to meet student needs and support proficient student work” (p. 10). To explicate further the Classroom Evaluation/Assessment standard, KDE used the following eight indicators to assist districts and schools in identifying best practices:

2.1a – Classroom assessments of student learning are frequent, rigorous and aligned with Kentucky’s core content.

2.1b – Teachers collaborate in the design of authentic assessment tasks aligned with core content subject matter.

2.1c – Students can articulate the academic expectations in each class and know what is required to be proficient.

2.1d – Test scores are used to identify curriculum gaps.

2.1e – Multiple assessments are specifically designed to provide meaningful feedback on student learning for instructional purposes.

2.1f – Performance standards are clearly communicated, evident in classrooms and observable in student work.

2.1g – Implementation of the state-required Assessment and Accountability Program is coordinated by school and district leadership.

2.1h – Samples of student work are analyzed to inform instruction, revise curriculum and pedagogy, and obtain information on student progress. (KDE, 2004b, pp. 10-17).

Stiggins (2008) defined assessment as a system of gathering information to drive instructional decision making (p. 1). Subsequently, Stiggins listed the following questions to guide assessment development: (a) What decisions? (b) Who's making them? and (c) What information will be helpful to them? (pp. 1-2). Balanced assessment systems provided opportunities for essential information to be disseminated at the classroom, school, and district levels (p. 2). In addition, in the Stiggins research on assessment accountability, each level of the educational organization is designed to address its own respective assessment-related question: (a) classrooms answer, Where is each student on the journey toward standards mastery? (b) schools answer, Which of our programs need improvement? and (c) districts answer, Are enough students achieving standards mastery?

(p. 2). While those involved at each level must answer its respective question on assessment accountability, no level can be expected to assist the others in “picking up slack” for prior or later levels’ failure to answer the respective assessment-related questions adequately (p. 2).

In developing a system of accountability within the reform framework of the 1990s, KDE outlined an intention for classroom assessment to move from a summative assessment construct to a formative model. KDE included specific language in Indicator 2.1a in Standard 2, Classroom Evaluation/Assessment, such as “frequent, rigorous, and aligned,” that parallels the descriptors of formative assessment examples (2004b, p. 10). In contrast, the summative assessment model was predicated on summative tests, whereby students had one opportunity to prove mastery of the content. While summative assessments provided valuable information on the overall learning taking place in the classroom upon the completion of an extended instructional period (e.g., after a unit of study, end of a nine weeks grading period), these assessments were only one measure for classrooms that demanded continuous feedback from both teacher and learners in mastering content (Stiggins, 2008). Stiggins declared “while such testing [standardized achievement tests] serves valuable accountable purposes, literally decades of obsessive belief in and reliance on such assessments has revealed that they cannot do the job” (p. 1).

The seminal research on formative assessments as a means of enhanced student achievement seemingly coincided with the passage of KRS 158.6455 by the Kentucky General Assembly in 1998—the legislation prompting the SISI whole school reform document (Clarke, 2008). During the late 1990s, Black and Williams confirmed formative assessments as a highly effective technique for enhancing student achievement and attainment, in addition to promoting lifelong learning (as cited in Clarke, 2008, p. 8). The

National Research Council (2000) defined formative assessment as the continual process of evaluating student learning to make thinking visible to both teachers and students (p. 24). Formative assessments assist teachers in gauging where students are in the “developmental corridor” from informal to formal thinking and to make subsequent adjustments based on this information (p. 24). Eschewing the typical “test” quality of more formalized summative testing, formative assessments are “learner-friendly” and serve as a roadmap for future instruction (pp. 24-25). Reeves compared the differences between a summative and formative assessment to an autopsy and physical examination, respectively (as cited in DuFour et al., 2012, p. 55).

In examining the current research on the types of assessment, Clarke (2008) described three strands: (a) assessment of learning, (b) assessment for learning, and (c) assessment as learning (p. 9). In terms of assessment of learning, this type of assessment is defined as a summative means of acquiring information, whether class-based, school-based, or a national test (p. 9). Regarding assessment for learning, these assessments are characterized as a feedback process that affords students the information about what to do to improve the assessed skill or mastery of standard (p. 9). And, when depicting assessment as learning, these types of assessments are defined as measures which assist students in transitioning from “what to improve” to “how to improve” (p. 9). In thinking about classroom evaluations and assessments, KDE (2004b) provided a foundation of evidence and descriptive language to promote assessment for learning and even assessment as learning, as described in Indicator 2.1e with its language for teachers to “provide meaningful feedback on student learning for instructional purposes” (p. 14).

In addition to increasing the frequency of rigorous, aligned assessments, KDE determined that schools should also be engaged in active collaboration among teachers for

the planning and design of authentic assessments—as described by KDE (2004b) in Indicator 2.1b (p. 11). Within the PLC module of DuFour et al. (2006), collaboration around results was accomplished through the following methods: (a) common preparation time; (b) parallel scheduling; (c) adjusted start and end times; (d) sharing classes; (e) group activities, events, and testing; (f) banked time throughout the school calendar; and (g) in-service and faculty meeting time (p. 97). KDE (2004b) noted in Indicator 2.1b for Standard 2, and specifically the Level 3 performance rating, the link to school leadership honoring this collaboration time through effective and frequent feedback to teachers on the development of assessment tasks (p. 11).

Additionally, KDE (2004b) embedded in Indicator 2.1b the concept of “authentic” assessment practices (p. 11). Moving away from the strict paper and pencil tests, KDE recommended assessment tasks such as exhibits, videos, and story boards (p. 11). Wiggins (1990) defined authentic assessment as a direct examination of student performance on “worthy intellectual tasks” (p. 2). In essence, authentic tasks match the assessment with the learning the student is required to master, from constructing an argumentative stance through an oratorical debate to developing an economics concept through a school fundraiser. With authentic assessments, teachers moved from the “objective, one-right-answer” type tests toward a learning and feedback process that informed student improvement on an authentic task. Through an opportunity to plan, improve, and modify learning on tasks, authentic assessment provides a valid and reliable measure by “emphasizing and standardizing the appropriate criteria for scoring such (varied) products” (p. 2). Through authentic assessments, students and teachers are afforded the opportunity to provide “meaningful feedback” toward student mastery, a component of Indicator 2.1e (KDE, 2004b, p. 14). According to Stiggins (2012), effective feedback was characterized

through the following attributes: (a) provides information about the work, not the learner; (b) delivers “actionable” strategies in a timely manner; (c) informs future learning modifications through descriptive language; and (d) gives clear instruction to the learner (p. 4). Through aligning authentic assessments to performance standards, and the subsequent provisions of feedback opportunities, Kentucky teachers were given a framework for moving from “teaching” to “learning.”

In addition to the development of rigorous assessments aligned to Kentucky standards, KDE (2004b) also emphasized the importance for student awareness of proficiency levels for these academic expectations through Indicator 2.1c (p. 12). In essence, where previous generations of assessments focused on what a student knew about a content standard, KDE stressed through Indicator 2.1c the requisite knowledge the student must display respective to a specific standard—with the underlying assumption that student mastery of a standard was reflected in the clear articulation of the standard by the student. In the level three performance rating for Indicator 2.1c, KDE included the process by which teachers collaboratively develop proficiency measures for each standard in each content area.

Collaborative development and reflection by teachers on what proficiency looked like was an important first step toward a uniform vision of proficiency for all educators in a school building and, subsequently, students. DuFour et al. (2006) offered two guiding questions for educators to consider when embarking on developing clear academic expectations for students: (a) What is it we want our students to learn? And (b) How will we know when each student has learned it? (p. 46). According to DuFour et al., during PLC collaboration to inform curricular and assessment decisions, the two aforementioned questions should be matched with the national, state, and local standards, aligned with

agreed upon proficiency levels, and assessed formatively to drive future modifications (pp. 46-47). In noting the intent of formative assessments to inform future instruction, the research of DuFour et al. effectively coincided with the KDE (2004b) Indicator 2.1d, which encourages educators to identify curriculum gaps through assessment scores (p. 13). Additionally, as described in Indicator 2.1c, through the clear articulation of academic expectations, first by teachers and then by students, performance standards can then be properly developed by teachers, displayed in the classroom learning environment, and evident in student work, as noted in Indicator 2.1f (p. 12, 15).

When examining Standard 2, Classroom Evaluation/Assessment, the final measure, Indicator 2.1h, provided by KDE (2004b) to inform school reform, was a powerful measure in embedding reflection components in Kentucky accountability and assessments. Through the analysis of student samples of work, as noted by KDE in Indicator 2.1h, groups of educators were afforded the opportunity to engage in a reflective practice to identify student proficiency. Additionally, Knight (2011) discussed the next step in moving reflection forward with subsequent action, based on the insight and evidence of reflection, and termed this protocol “praxis” (p. 20). Encouraging teacher modifications and revisions of curriculum and pedagogy to meet the needs of the learner, Kentucky educators began the important partnership principle of praxis. As evidenced by the SISI document, through the Academic Performance standard of Classroom Evaluation/Assessment and its respective indicators, accountability and assessment was moving toward a culture where teaching was being measured not by what was “covered” but by the learning taking place in the classroom.

Instruction

Standard 3, Instruction, of the SISI document comprised the third and final

standard of the Academic Performance group. KDE (2004b) defined Standard 3, Instruction, as “the school’s instructional program actively engages all students by using effective, varied, and research-based practices to improve student academic performance” (p. 18). Standard 3, Instruction, comprised of the following eight indicators, described the best practice evidence that KDE suggested for school improvement:

3.1a – There is evidence that effective and varied instructional strategies are used in all classrooms.

3.1b – Instructional strategies and learning activities are aligned with the district, school and state learning goals, and assessment expectations for student learning.

3.1c – Instructional strategies and activities are consistently monitored and aligned with the changing needs of a diverse student population to ensure various learning approaches and learning styles are addressed.

3.1d – Teachers demonstrate the content knowledge necessary to challenge and motivate students to high levels of learning.

3.1e – There is evidence that teachers incorporate the use of technology in their classrooms.

3.1f – Instructional resources (textbooks, supplemental reading, technology) are sufficient to effectively deliver the curriculum.

3.1g – Teachers examine and discuss student work collaboratively and use this information to inform their practice.

3.1h – There is evidence that homework is frequent and monitored and tied to instructional practice. (KDE, 2004b, pp. 18-25).

Through an examination of the indicators for Standard 3, Instruction, the first three indicators reflected a KDE (2004b) vision that schools focus on pedagogy grounded in best

practices, aligned to school, district, and state learning goals, and provided multiple avenues for mastery accessible to all learners (pp. 18-20). While some neo-conservative educational groups (Innes, 2010) criticized the ubiquity of the phrase “the research says...” during Kentucky reform of the 1990s, a considerable amount of time and research however had been conducted throughout the late 20th and early 21st century to gauge and confirm the strategies that had the greatest effect size on student growth and achievement (Hattie, 2009).

Beginning research on effective teaching and learning strategies in education in 1993, Zemelman, Daniels, and Hyde (2012) borrowed the term “best practice” for education from the medical community to describe “solid, reputable, state-of-the-art work in a field” (p. 1). Moving from educational assumptions to educational facts, Zemelman et al. grounded the best practices work in the abundant research on successful teaching strategies conducted by a variety of respected academic associations (National Board for Professional Teaching Standards, National Council of Teachers of Mathematics, American Association for the Advancement of Science, National Research Council, etc.). Through a synthesis of analyses on educational teaching strategies, Zemelman et al. aggregated the available research on effective instructional pedagogy into a condensed list of the following seven best practices for teaching: (a) gradual release of responsibility, (b) classroom workshop, (c) strategic thinking, (d) collaborative activities, (e) integrative units, (f) representing to learn, and (g) formative-reflective assessment (pp. 39-82).

In examining performance rating three (“fully functioning and operational level of development and implementation”) of Indicator 3.1a, KDE (2004b) cited as evidence of effective instruction the implementation of “student-centered, culturally responsive instructional strategies” such as cooperative learning, activity stations, and hands-on

learning (p. 18). Zemelman et al. (2012) also noted the advantages of incorporating these student-centered instructional strategies in the classroom, under the best practice structure for teaching, “classroom workshop” (pp. 46-49). Classroom workshop models encouraged student-centered learning as students directed their own learning with a selection of high interest topics for reading and writing (p. 46). Through classroom workshops, students also engaged in collaborative learning opportunities, recorded their own learnings and insights, and reflected on reading and writing through self-evaluation (p. 46). In their work on personalized learning, Bray and McClaskey (2015) cited students driving their own learning as an approach to instructional pedagogy that both motivated and challenged learners—and often forced students to work harder than the teacher (p. 11). Additionally, through the best practice structure of “collaborative activities,” Zemelman et al. (2012) stressed the value of decentralized groupings of students—partners, small groups, long-term teams (p. 57). Zemelman et al. cited landmark studies in the 1990s and 2000s that documented dramatic student achievement gains when students engaged in cooperative learning tasks (p. 57).

In addition to student-centered, cooperative instruction, KDE (2004b) proffered higher-order thinking skills and problem-solving activities as another requisite descriptor for an accelerated performance rating with Indicator 3.1a (p. 18). Zemelman et al. (2012) discussed the high level of thinking skills embedded in the best practice teaching structure, “strategic thinking” (pp. 50-57). However, Zemelman et al. stated that very little classroom instruction actually engaged students in high level thinking activities (p. 50). Pearson and Gallagher (cited in Zemelman et al., pp. 50-51) proposed that teachers model specific thinking strategies to students who are engaged in reading, an activity embedded in every content area whereby students: (a) monitor their comprehension; (b) make their

own sensory images; (c) make connections with their background knowledge; (d) ask themselves questions; (e) draw their own inferences; (f) determine their own importance; and (g) synthesize their own meaning. Teacher modeling of think-alouds to students was determined to be an extremely powerful process in promoting higher-order, strategic thinking by students (p. 51). In addition to helping students to realize even educated adults must use “careful, recursive, stepwise work” to break apart text, modeling thinking strategies enabled students to identify the thinking processes embedded in real-world activities such as scientists forming hypotheses (p. 51). Additionally, Hattie (2009) listed problem-solving teaching as having a powerful effect size of .61 on student achievement, where $d = .40$ was the typical impact of a given educational influence and $d = .61$ was classified a high impact strategy in the zone of desired effects (p. 17, p. 297).

According to Ogawa, Sandholtz, Martinez-Flores, and Scribner (2003), standards-based reform measures, much like the standard-alignment measures reflected in indicator 3.1b of SISI Standard 3, Instruction, impacted teacher instruction in the following three areas: (a) greater consistency among teachers at each grade level as standards were implemented to assess instructional goals; (b) revisions made in curriculum to align with standards, sometimes eliminating unassessed content areas; and (c) organizational changes implemented during instructional time to focus on assessed standards (pp. 165-166).

Ogawa et al. stated that standards provided a cogent framework for instruction to replace the former loosely organized curriculum of many schools and districts (p. 147).

Within the SISI framework for school improvement which emphasized the strategies for enhancing learning and student outcomes, KDE (2004b) additionally highlighted the integral importance of school leadership in monitoring and assessing best practices by teachers (p. 20). Indicator 3.1c outlined measures for administrative teams to

ensure that best practices were implemented, with strategies that met the various needs of a diverse population through brain-based research and multiple intelligences (p. 20). The work of Gardner (1983) on multiple intelligences, which eschewed the common “fixed IQ intelligence” beliefs of many educators, encouraged teachers to access pathways to enhancing student learning through the following eight multiple intelligences: (a) linguistic intelligence, (b) logical-mathematical intelligence, (c) spatial intelligence, (d) bodily-kinesthetic intelligence, (e) musical intelligence, (f) interpersonal intelligence, (g) intrapersonal intelligence, and (h) naturalist intelligence. Many years after the seminal work on multiple intelligences, Gardner warned against the skewing or misinterpreting his research since it was essentially grounded on two main principles: (a) have an awareness of differences among individuals in order to focus on the essential elements of the discipline and (b) teach these essential components in a variety of ways (as cited in Hattie, 2012, p. 91). Hattie (2012) stated that focusing on the two aforementioned assertions by Gardner aligned with the higher effect size strategy of encouraging “multiple ways of teaching” (p. 91).

Indicator 3.1d for SISI Standard 3, Instruction, included as descriptors of effective schools the administrative recruitment and development of highly qualified teachers with appropriate certification, respective to their content and grade level (KDE, 2004b, p. 21). Multiple studies have confirmed the overwhelming value that an effective teacher has on the growth and achievement of students (Hattie, 2009, 2012). Zuckerman stated that students who have an effective teacher for three consecutive years scored 50% higher on achievement measures than their counterparts (as cited in Zemelman et al., 2012, p. x). Additionally, KDE (2004b) cited continuous professional development of teachers in challenging and inspiring students as another characteristic of schools with highly effective

instructional programs (p. 21). Bausmith and Barry stated that effective features of a quality professional development program included coaching teachers, implementing school-wide data teams, focusing on how students learn in the respective content areas, and collaborating to plan and assess student learning recursively based on data (as cited in Zemelman et al., 2012, p. 175). Additionally, KDE (2004b) suggested this same professional endeavor in Indicator 3.1g whereby teachers continuously focused and examined student work samples, with a subsequent process for identifying gaps in student learning to drive future instruction (p. 24). Zemelman et al. (2012) proffered “formative-reflective assessment” as another best practice teaching structure for high impact pedagogy, whereby teachers examined and gauged student work not just for grade justification, but for making essential and intentional modifications to future instruction (pp. 78-79).

In an effort to support and enhance instructional pedagogy, KDE (2004b) included Indicator 3.1e and 3.1f to encourage the use of technology and instructional resources, respectively, in the classroom (pp. 22-23). For Indicator 3.1e, regarding the evidence required for a level three performance rating, KDE cited the use of technology as a conduit to extend and enrich learning opportunities for students and as a means to create a multitude of products for varying authentic audiences (p. 22). In thinking about how technology promoted enhanced learning, Magaña and Marzano (2014) cautioned against using classroom technology in isolation of research-based pedagogy: “technology, it seems, cannot yield achievement gains in a vacuum” (p. 19). Only through the knowledge and application of high impact instructional strategies by teachers, in confluence with appropriate technological software and hardware, could learning outcomes truly be maximized (p. 1). Additionally, KDE (2004b) cited as evidence of effective pedagogy, in

terms of instructional technology, that teachers and classrooms encouraged opportunities for student connections to a greater civic society (p. 22). Current practices for technological collaboration include authentic audiences for students through Project-Based Learning, linking students to real world audiences within the community and across the globe. In addition to embedding technology in instructional practices, KDE (2004b) noted in Indicator 3.1f the inclusion of a variety of “sufficient” instructional resources to bolster and enhance the curriculum across the content areas (p. 23). In aligning the instructional resources to the content and standards of Kentucky education, educators were encouraged to monitor the implementation of current, research based instructional resources continually in order to respect the diversity of learners, individual learning styles, and enhance educational outcomes (p. 23).

Through the final example of instruction for whole school reform, KDE (2004b) listed homework in Indicator 3.1h as an effective pedagogical structure when “frequent and monitored and tied to instructional practice” (p. 25). In further describing homework that reached the level of the performance three rating for the Scholastic Audit based on the SISI document, KDE emphasized the importance of students being able to express the purpose of homework and how it connected to classroom instructional practices (p. 25). However, through a meta-analysis of the influence of homework on student achievement during the past twenty years, Cooper, Robinson, and Patall (as cited in Hattie, 2012, p. 12) noted homework as having an average effect size ($d = .40$) across all grade levels, with a negative impact on student achievement respective to elementary schools ($d = -.08$). Further synthesis by Hattie across five meta-analyses confirmed the low to average impact of homework on student achievement with an effect size of $d = .29$, when compared to the ‘hinge point’ effect size for an educational influence of $d = .40$ (p. 13). Nevertheless, the

inclusion of KDE (2004b) in prompting educators to provide specific, timely feedback for all homework assignments, actually had a tremendous impact on student achievement when applied to all elements of learning (Hattie, 2009). Hattie highlighted feedback as one of the most influential factors in enhancing student achievement, with a high effect size of $d = .73$ in the “zone of desired effects” (p. 173). Hattie and Timperley (as cited in Hattie, 2009, p. 174) defined feedback as “information provided by an agent (e.g., teacher, peer, book, parent, or one’s own experience) about aspects of one’s performance or understanding.” In describing feedback through a variety of avenues, not strictly “teacher to student” feedback, Hattie understood the tremendous impact that feedback had not only on homework, but as a “consequence of performance” that could be gauged through a multitude of performance measures, in measuring student achievement (pp. 173-174).

Through the Academic Performance standards of the SISI document (Standard 1, Curriculum; Standard 2, Classroom Evaluation/Assessment; Standard 3, Instruction), KDE comprehensively grounded efforts for whole school reform through the aforementioned classroom practices that directly impacted student learning and achievement. In examining each SISI standard through a list of indicators that specified best practices, Kentucky educators and Scholastic Audit and Review teams had a valid, reliable tool through which to measure performance in the classroom. The following section in the current review of literature examines the state of curriculum, classroom evaluation and assessment, and instruction through the lens of elementary science.

Elementary Science

Within the current study, the researcher focused analysis on the elementary level in assessing the impact of Academic Performance from the first three standards of the *Standards and Indicators for School Improvement (SISI)* on science achievement scores.

As previously referenced in Chapter I, the specific selection of elementary level schools as the unit of analysis was an intentional decision by the researcher, predicated on the following rationale: First, no current studies have assessed the impact of the Academic Performance standards from the SISI specifically or generally on science at the elementary level. Second, in response to declining international science achievement scores for the United States (National Research Council, 2010; U.S. Department of Education, 2015), an examination of effective classroom academic practices at the elementary level was warranted—since elementary science provided the scientific foundation of knowledge and application that is later expanded and assessed at the middle and high school level. Third, since Kentucky elementary school teachers are content generalists, as opposed to their content-specific secondary level counterparts, an examination of the impact of Academic Performance standards from the SISI on elementary science achievement could help determine whether exemplar instructional practices transcend a single, respective content area. Fourth, from the varying extant literature on achievement gaps (Evans, 2005), students are observed to suffer a widening achievement gap for a variety of school, cultural, and socioeconomic-based factors while progressing through the traditional kindergarten to twelfth grade levels; thus, perhaps the current research could shed light on best practices specifically in science that if implemented would play a role in the reduction of gaps in earlier grades such as elementary school.

Taking into consideration the aforementioned importance of science at the elementary level, this section examines the classroom best practices of the Academic Performance standards in the SISI (Standard 1, Curriculum; Standard 2, Classroom Evaluation/Assessment; and, Standard 3, Instruction) and focuses on the application of these practices through the specific lens of elementary science. Additionally, some of the

following sections on elementary science address advances in the field—specifically, the Next Generation Science Standards (NGSS)—and their relationship to best practices to the SISI Standard 1, Curriculum.

Science Curriculum

Among the Academic Performance standards for the SISI, the first standard in order, and perhaps importance, was Standard 1, Curriculum. Marzano (2003) listed five school level factors that impacted student achievement: (a) guaranteed and viable curriculum; (b) challenging goals and effective feedback; (c) parent and community involvement; (d) safe and orderly environment; and (e) collegiality and professionalism (p. 15). Marzano continued by citing “guaranteed and viable curriculum” as the most influential school-level factor on student learning (pp. 15, 22). Marzano bifurcated a “guaranteed and viable curriculum” into the two respective components of “opportunity to learn” and “time” (p. 22).

In terms of an opportunity to learn for curriculum, Marzano (2003) summarized this concept as students having ample opportunity to learn the required content (p. 24). A guaranteed curriculum required both district/school and classroom/teacher commitments: clarity and verticality of content among grade levels and an intentional focus on the content by classroom teachers (p. 24). According to Marzano (2013), viability, the second equally important factor included in the school-level factor of a “guaranteed and viable curriculum,” underscored and was a requisite element for a guaranteed curriculum (p. 38). In order for a curriculum to be viable, school teachers must have adequate time during the instructional school year to address the specific content (p. 38). Interestingly enough, Marzano cited the lists of content-specific standards, developed by respective content-specific educational advocacy groups of the 1990s, as saturating classrooms with so much

content that the curriculum could not reasonably be addressed by the teacher during the school year—making it neither viable nor subsequently guaranteed (p. 38). In fact, Hinde (2005) also decried the availability of adequate time to teach all the different content standards sufficiently: “it is not feasible to expect teachers to address all the mandated standards in the course of a school year” (p. 105). Hinde continued an indictment on the inability to teach the entire curriculum by citing the heavy time demands of reading and literacy in the elementary school, specifically the primary grades, that precluded teaching content standards of other subjects (p. 105). Invoking education philosopher John Dewey, Hinde offered the following:

The reality of today’s elementary classrooms, however, are such that teachers are pressured to devote most of their time and energy on areas that are tested and to avoid considering what is being lost by the narrowing of the curriculum—the appreciation of things worthwhile, the values to which those things are relative, the desire to apply learning, and the ability to extract meaning from future experiences. (p. 105)

Marzano (2003) continued his research into what impacts student achievement by amassing all the studies on student learning and then listing the following three high impact factors: (a) instructional strategies, (b) classroom management, and (c) classroom curriculum design (p. 71). When specifically discussing curricular measures, Marzano espoused that classroom curriculum design included intentional decisions made by the individual teacher, even with a “guaranteed and viable curriculum” in place at the district and school level (p. 71). While perhaps the least analyzed of the three teacher-level factors, Marzano urged educators not to overlook classroom curriculum design as it provides a strong effect size for student achievement (p. 106). In essence, effect size

gauges the impact of a sample treatment from the control treatment, with the difference divided by the standard deviation. With an extensive literature research base, Marzano found a wealth of pragmatic structures that could be implemented and embedded into the classroom curriculum (p. 106). Also, Marzano noted that many failures in student ability to master content could be connected to lackluster classroom curriculum design (p. 106). Marzano defined the phrase “classroom curriculum design” as the “sequence and pacing of content along with the experiences students have with that content” (p. 106). Of particular note, Marzano continued by emphasizing the term “classroom” in the phrase “classroom curriculum design” since the individual classroom teacher possesses a great deal of autonomy regarding when the content would be taught, how long the content would be taught, and how the content would be taught (p. 106).

Marzano lamented the reliance of classroom teachers on textbooks in pacing the content, noting a Rothman study that depicted science textbooks as “well-illustrated dictionaries as opposed to effective vehicles for student learning” (as cited in Marzano, 2003, p. 107). Eschewing some of the cognitive strategies that are counteractive if overused to grounded classroom curriculum design, such as “constructivism” and “brain-based education,” Marzano offered the following principles for educators in curriculum design at the classroom level to enhance student learning: (a) identifying a focus of specific types of knowledge for the unit or lesson; (b) engaging in structured tasks that afford opportunities for effective acquisition of these specific types of knowledge; and (c) providing multiple opportunities for complex interactions with these specific types of knowledge (pp. 109-116).

Addressing the aforementioned deficit of time to teach all the content standards sufficiently and incorporating many of the classroom curriculum design elements above,

many elementary teachers employ the “best practice” of curriculum integration through the development of extended, intensive learning units that incorporate all subjects—i.e., an Ocean unit, a Seasons unit, etc. (Marzano, 2003; Zemelman et al., 2012, p. 68).

Additionally, thoughtful curriculum integration provides various opportunities for students to engage with the content from several different subjects and several different angles. For example, Zemelman et al. stressed the importance of how teachers can organize and re-adjust the content to make it more appealing and meaningful for students—with the location of underlying themes and emphasizing the “big ideas” of the text (p. 69).

Over the past 50 years, the goals, directives, and assumptions of elementary science curriculum have progressed insignificantly, in some instances, and dramatically, in other occurrences (Sandall, 2003). During the 1960s, curriculum development in elementary science focused on two main assumptions: (a) science presented through the lens of the science profession would appeal to students of all ages and (b) any subject can be taught to any child, regardless of age (p. 14). Nonetheless, according to Harms and Yager (1981), elementary science during the 1970s looked drastically different with science only being taught at the end of the day and only if there was enough time—by ill-equipped, ill-prepared content-generalists. By the 1980s, Staver and Bay (1987) found that educators were more concerned with scientific and technical literacy for students than other avenues of elementary science instruction. With an emphasis on scientific and technical literacy among other factors, Campbell, Voelkl, and Donahue (1996) recognized an increase in the National Assessment of Educational Progress (NAEP) in all levels of performance from 1977-1996.

As mentioned previously, Kentucky developed the *Core Content for Assessment* (KDE, 1996) during the mid-1990s to guide districts and schools in teaching the core

subjects (Petrosko, 2000; Petrosko et al., 2000). In the section below, the researcher specifically discusses the *Core Content for Assessment* (KDE, 1996) in science.

Kentucky science core content. Representing the high point of Kentucky education reform, the *Kentucky's Learning Goals and Academic Expectations* (KDE, 2012c) outlined the following science knowledge and skills students should know and demonstrate respectively, following their Kentucky school experience:

Learning Goal 2: Students shall develop their abilities to apply core concepts and principles from mathematics, the science, the arts, the humanities, social studies, practical living studies, and vocational studies to what they will encounter throughout their lives.

Academic Expectations for Science:

2.1 Students understand scientific ways of thinking and working and use those methods to solve real-life problems.

2.2 Students identify, analyze, and use patterns such as cycles and trends to understand past and present events and predict possible future events.

2.3 Students identify and analyze systems and the ways their components work together or affect each other.

2.4 Students use the concept of scale and scientific models to explain the organization and functioning of living and nonliving things and predict other characteristics that might be observed.

2.5 Students understand that under certain conditions nature tends to remain the same or move toward a balance.

2.6 Students understand how living and nonliving things change over time and the factors that influence the changes. (pp. 1-2)

The Kentucky *Core Content for Science Assessment*, Version 4.1 (KDE, 2006b), and previous iterations such as Versions 3.0 (2000a) and 4.0 (2005a), constituted a subset of *Kentucky's Program of Studies for Grades Primary-12* (KDE, 2006c). While not an exhaustive list of the science knowledge for students, the *Core Content for Science Assessment*, Version 4.1 (KDE, 2006b), included the most essential science content, excerpted from the more comprehensive *Kentucky's Program of Studies for Grades Primary-12* (KDE, 2006c) and *Kentucky's Learning Goals and Academic Expectations* (KDE, 2015b), which would then be eligible for assessment of Kentucky students (2006b, p. 2). According to Karen Kidwell, a Director for KDE (personal communication, March 15, 2016), the *Core Content for Assessment*, Version 4.1 (KDE, 2006b) addressed only a subset of standards. KDE warned in its introduction to the *Core Content for Science Assessment*, Version 4.1, that the content included within the respective document was not the overall and comprehensive local curriculum to be used for science, nor was it the minimum expectations required for elementary, middle, and high school level programs that contributed to graduation requirements (p. 2).

Petrosko (2000) cited the *Core Content for Assessment* (KDE, 1996) as the “most specific curriculum guidance document” by KDE, in terms of classroom instruction (Petrosko, 2000, p. 28). According to teachers interviewed by Petrosko, the *Core Content* was predominantly used for the alignment and planning of curriculum (p. 43). Additionally, science topics were usually divided up among teachers for subsequent instruction (p. 43). Science texts were aligned with national standards and, thus, also Kentucky in the *Core Content for Assessment* (as cited in Petrosko, 2000, p. 43).

For the current research, all elementary schools were assessed with the Kentucky *Core Content for Science Assessment*, Version 3.0 (KDE, 2000a), as the primary science

document of content standards during the 2001-2005 testing cycles. Kentucky *Core Content for Science Assessment*, Version 3.0, contained three different sections: (a) Conceptual Understandings of Physical, Life, and Earth/Space Science; (b) Scientific Inquiry; and (c) Applications/Connections, the understanding of the nature and utility of science (p. 1). Furthermore, in terms of Scientific Inquiry and Applications/Connections, these sections of science were assessed in terms of the three aforementioned divisions of the first section—Physical, Life, and Earth/Space Science (p. 1). Test items gauged student understanding not merely of facts and details, but, furthermore, of the “ability to reason and use problem-solving skills developed through inquiry and the application of scientific concepts to real-life situations” (p. 1). While the Kentucky test of science core content became progressively more abstract for students in middle and high school, elementary students were assessed on their understanding of concrete concepts and the relationships among these aforementioned concepts (p. 1). In the following paragraphs, the researcher briefly describes each of the three sections of the Kentucky *Core Content for Science Assessment*, Version 3.0.

First, in terms of Conceptual Understanding, the term *understanding* referred to the broader categories of comprehension that included concepts, relationships of concepts, reasons for relationships, how to use these ideas behind the relationships, and subsequent methods for applying these ideas (KDE, 2000a, p. 1). Furthermore, content statements under the section on Conceptual Understanding delineated an emphasis on “science facts, concepts, principles, theories, and models that are important for all students to know, understand, and use” (p. 1).

Second, under the Scientific Inquiry section of the *Core Content for Assessment*, Version 3.0 (KDE, 2000a), KDE defined scientific inquiry as the various ways scientists

observe the natural world and propose rationale for different phenomena (p. 1). Scientific inquiry included a variety of active processes for exploring the world: observing, questioning, examining, analyzing, and interpreting. Subsequent proposals of hypotheses and explanations, and communicating these results to others, framed the scientific inquiry foundation that was to be assessed (p. 1).

Third, under the section of Applications/Connections, KDE (2000a) divided this concept into three sub-categories: science and technology, with its applications for effective and efficient science practice; personal and social perspectives, with a study of population, natural resources, the environment, and both artificial and natural hazards; and the history and nature of science, with a focus on the historical progression of science through reason, the qualities of science as knowledge, and historical perspectives (p. 1). Students were assessed with test items regarding Applications/Connections the advancement of science through technology, the relationships among human perspective and scientific challenges, and the human influence on science and the role of science in the world (pp. 1-2).

In the following section, the researcher briefly discusses the progression of curriculum in Kentucky, with its evolution toward the eventual adoption of national standards, and the subsequent impact on what is taught and assessed at the school and classroom level.

National Next Generation Science Standards. The *Next Generation Science Standards* (NGSS) document was developed in a two-step process by the National Research Council (NRC), the National Science Teachers Association (NSTA), the American Association for the Advancement of Science, and Achieve (as cited in “Development Overview,” 2016, p. 1). Step one included “getting the science right” as the

NRC composed *A Framework for K-12 Science Education* (2012). Through a committee of 18 members esteemed in different fields of science, the NRC embedded the most current research on science learning into a comprehensive list of science standards that K-12 students should know and understand (“Development Overview,” 2016, p. 1). Step two included an Achieve-led facilitation of state input for the NGSS development (p. 1). In fact, Kentucky was one of the lead states to provide input for the NGSS. With the initial insight of *A Framework for K-12 Science Education* (NRC, 2012), the NGSS reflected K-12 science standards that were substantive and coherent and provided science knowledge across science disciplines and domains to all students with an internationally benchmarked education (“Development Overview,” 2016, p. 1). Additionally, the NRC (2012) conducted a fidelity review of the final NGSS draft that confirmed its alignment with the *Framework for K-12 Science Education* predecessor (“Development Overview,” 2016, p. 1).

The NGSS was developed and implemented (awaiting implementation in Kentucky as of 2016) after the current Scholastic Audit data were collected and was thus beyond the boundaries of this study (as cited in National Science Teachers Association, 2013, p. 1). However, a brief description and analysis of the NGSS is included in the following paragraphs to provide a comparison to the aforementioned *Kentucky Core Content for Science Assessment*, Version 4.1 (2006b).

In essence, the NGSS (NGSS Lead States, 2013) outlined a new vision for science education in America (National Science Teachers Association, 2013, p. 1). Moving from the way things are to the way things should be, the NGSS outlined the following seven conceptual shifts to be implemented in science education, detailed in subsequent paragraphs:

Shift 1: K-12 science education should reflect the interconnected nature of science as it is practiced and experienced in the real world.

Shift 2: The NGSS are student performance expectations, not curriculum.

Shift 3: The science concepts in the NGSS build coherently from kindergarten through 12th grade.

Shift 4: The NGSS focus on deeper understanding of content as well as application of content.

Shift 5: Science and engineering are integrated in the NGSS, from kindergarten through 12th grade.

Shift 6: The NGSS is designed to prepare students for college, career, and citizenship.

Shift 7: The NGSS and Common Core State Standards (English language arts and mathematics) are aligned. (pp. 1-3)

In terms of the first conceptual shift, the NGSS required educators to embark in instruction at the nexus of three dimensions: science and engineering practices, crosscutting concepts, and disciplinary core ideas (p. 2). With present emphasis placed separately on each of the aforementioned dimensions, it will take wholesale reform to all aspects of science—instruction, curriculum, assessment, teacher preparation, and professional learning—in order to align with this NGSS vision.

With the second conceptual shift, the NGSS included standards that combined performance expectations and foundation boxes among the three dimensions, rather than a prescriptive, step-by-step guide on how to teach the standards and what constituted the curriculum (p. 2). The third conceptual shift noted the change to the scaffolded vision of the *Framework for K-12 Science Education*, through its notion of science knowledge as an

evolution of many years, rather than a single unit during a school year. In terms of the fourth conceptual shift of the NGSS, an emphasis is placed on the depth—not breadth—of science knowledge. The NGSS now moved educators from a list of facts that constituted a science education to an in-depth analysis of core ideas.

For the fifth shift in conceptual understanding of science education, the developers of the NGSS stressed the importance of technology and engineering in a thorough understanding of science by weaving these two constructs throughout the NGSS (p. 2). The sixth conceptual shift of the NGSS incorporated this component of college and career readiness, in addition to enhanced citizenship, for all students—regardless of future college and career plans or potential STEM field majors (p. 3). Finally, the seventh conceptual shift of the NGSS was grounded in its concerted effort to align with the Common Core State Standards in English language arts and mathematics. The NGSS works in tandem to promote integrated teaching and learning in all three fields, rather than in separate content areas.

Currently, as of 2016, Kentucky has adopted the NGSS, with several in whole—including climate change and evolution, once points of contention by Kentucky lawmakers. Additionally, Kentucky has just appointed a new Commissioner of Education, Dr. Stephen Pruitt, who served on one of the original NGSS development committees (personal communication, November 7, 2016). For the current research, the author focuses analysis on the Kentucky *Core Content for Assessment*, Version 3.0 (2000a), as the guiding curriculum document. In the following section, the researcher discusses best practices for SISI Standard 2, Classroom Evaluation/Assessment, in terms of elementary science.

Classroom Evaluation/Assessment in Science

As mentioned previously in the current review of literature, SISI Academic Performance Standard 2, Classroom Evaluation/Assessment, was defined by KDE (2004b) as the school or classroom using a variety of assessment practices for continuous improvement of student achievement in a content area (p. 10). In the content area of science, as observed through nearly a decade of researcher observation, classroom evaluation and assessment best practices include a multitude of measures to gauge student learning, such as the articulation of learning expectations and formative assessments. While other content areas assess students through traditional, pencil-and-paper based measures, the active nature of science demands a more hands-on type approach to evaluating student success.

Formative assessments are assessments *for* learning that occur periodically and intentionally throughout the learning process during a unit of instruction (DuFour et al., 2006, p. 55). DuFour et al. additionally cited formative assessments as a way for teachers to alleviate the pressures of state testing—not truly knowing student comprehension until the end of the year--through frequent monitoring of student learning (p. 49). Berger, Rugen, and Woodfin (2014) cited numerous studies that verified formative assessments as enhancing achievement, instruction, and motivation (p. 10). Black and Wiliam stated “firm evidence shows that formative assessment is an essential component of classroom work and that its development can raise standards of achievement” (as cited in Berger et al., 2014, p. 15). Additionally, Black and Wiliam ascertained that formative assessments nearly doubled the speed of learning (as cited in Berger et al., 2014, p. 10). Quality formative assessments align with SISI Standard 2, Classroom Evaluation/Assessment, Indicator 2.1a, whereby evaluations of student learning are “frequent, rigorous and aligned with Kentucky’s core content” (KDE, 2004b, p. 10).

Marzano and Brown (2009) championed formative assessments as both “powerful measurement tools and powerful instructional tools” (p. 25). In terms of a powerful measurement tool, formative assessments afford teachers multiple opportunities to gauge student learning before the typical, end-of-unit summative assessment, aligning accordingly with Indicator 2.1d for identifying gaps in the curriculum to inform future modifications (Marzano & Brown, 2009, p. 25; KDE, 2004b, p. 13). Additionally, formative assessments in science serve as a powerful instructional tool by providing students with detailed information on the exact types of content standards they have yet to master and subsequently ask for help or work harder to grasp concepts—again, before the typical, end-of-unit summative assessment (p. 25). Formative assessments as an instructional tool coincides with SISI Standard 2, Classroom Evaluation/Assessment, Indicator 2.1e, for providing valuable feedback to students regarding their learning (KDE, 2004b, p. 14). In essence, formative assessments are another tool of best practice for educators (National Research Council, 2007, p. 279). Shavelson and Stanford Education Assessment Laboratory found that formative assessments give both teachers and students agency to enhance learning continuously throughout a unit of instruction (as cited in National Research Council, 2007, p. 279). Understanding and monitoring ideas is at the core of scientific thinking and, consequently, formative assessments provide perhaps the best tool for monitoring and evaluating student thinking (p. 279).

Student articulation and reflection of expectations and objectives represent another best practice in science that aligns with SISI Standard 2, Classroom Evaluation/Assessment, Indicator 2.1c, whereby students are encouraged to “articulate academic expectations in each class and know what is required to be proficient” (KDE,

2004b, p. 12). According to the National Research Council (2007), articulation and reflection stand as two processes that reside at the heart of science learning:

In scientific practice, constructing and testing knowledge claims require a focus on articulating those claims, that is, developing clear statements of how and why phenomena occur. Argumentation requires articulating claims and teasing apart when there is agreement or divergence among different claims. (p. 278)

The National Research Council continued by citing reflection as another integral process in scientific discovery, especially when conducting an investigation (p. 278). Interestingly, the authors cited reflection not only as an end-of-experiment cognitive process, but, also, an “ongoing articulation of understanding and reflection on both the practices and the content of the investigation” (p. 279). Additionally, Berger et al. (2014) cited reflection as an important element of student-engaged assessments, one that should be embedded frequently throughout the classroom. At its most impactful, reflection should move beyond “vague statements of preferences, strengths, and weaknesses” with a focus on depth of thinking (p. 9).

Through the articulation of learning expectations and formative assessments, science educators are offered momentary glimpses inside the learning process of students. In the following section, the last SISI Academic Performance standard—Standard 3, Instruction—is addressed in terms of elementary science and best practices.

Science Instruction

The National Research Council (2007) noted, “regardless of one’s theoretical orientation, by the time children enter elementary school, no one would argue that their minds are empty vessels awaiting enlightenment in the form of instruction” (p. 53).

Examining the best practices research over the past two decades on science instruction, and

aligning these best practices to the eight indicators for SISI Standard 3, Instruction, several trends emerged whereby best practices afforded another layer of specificity and direction for elementary science educators.

According to the National Research Council (2007), there has been a dramatic rise in the understanding of learner development, and how students learn, over the past two decades (p. 19). The authors noted the influence of prior knowledge—both conceptions and misconceptions—in making sense of the natural world (p. 19). In terms of instruction, taking into consideration the impact of prior knowledge on learning, scientific reasoning and science content should not be taught to the exclusion of the other or in silos (p. 19). Additionally, new research over the past twenty years also facilitated an enhanced understanding of learning environments for science students. No longer are students confined to the textbook for scientific learning: “children learn science from books, television, the Internet, visits to museums and national parks, as well as the science classroom and the scientific and technological world around them” (p. 19). Furthermore, learners use knowledge and language to question, challenge, confirm, and explore the natural world. Thus, the National Research Council recognized this ongoing conversation about science among students, noting that peer discourse helped students establish an understanding of the scientific world (p. 19). Bybee (2002, p. 31) noted that science instruction possessed a complexity that, when conducted at an optimal level, combined students, content, and the physical learning environment, as teachers made long-term decisions about the curriculum and immediate decisions regarding learner understanding. Accordingly, the aforementioned, evidence-based research aligns with Indicator 3.1a of SISI Standard 3, Instruction, since best practices in learning must be understood to achieve the component of “effective and varied instructional strategies are used in all classrooms”

(KDE, 2004b, pp. 18).

The National Research Council (2007) identified three main themes of organizing research on how young students begin to develop an understanding of the scientific universe: (a) a concern with explanation and investigation for young learners is paramount and “there is no simple concrete to abstract progression in children’s development;” (b) explanatory insights emerge for children from the four specific domains of mechanics, folk biology, rudimentary chemistry, and folk psychology; and (c) an “interplay between domain-specific forms of learning and domain-general ones is central to any account of the emergence of scientific thought” for young learners (pp. 54-55). Bybee (2002, p. 31) cited children’s curiosity as a main motivator for making sense of the world. Bybee explored this curiosity, in both children and adults, and cited the following pattern of normal inquiry and formal problem solving, respectively: (a) initial engagement, (b) exploration of alternatives, (c) formation of an explanation, (d) use of the explanation, and (e) evaluation of the explanation based on its efficacy and responses from others.

For the remainder of the current section on Science Instruction, the researcher discusses general pedagogy in the elementary science classroom, and the two emerging trends regarding the “nature of science” and “inquiry-based learning.”

Pedagogy in the elementary classroom. Since the *Kentucky Education Reform Act of 1990* (KERA), developed in response to a lawsuit from 66 school districts detailing the inequities in Kentucky school finance allotment, Kentucky schools have been accountable for achieving high standards for all students (Kentucky Department of Education [KDE], 2008). The Kentucky Supreme Court during the landmark ruling in 1989, which prompted KERA, declared “each child, every child, in this Commonwealth must be provided with an equal opportunity to have an adequate education” (KDE, 2000b,

p. 10). For the first time in Kentucky education, public schools had a level-playing field in finance, remediation opportunities for at-risk children, and clarity of learner goals with high standards of accountability (KDE, 2000b).

While setting high standards is integral to school reform and improvement, the planning and implementation of exemplar curriculum, instruction, and assessment is perhaps even more important. According to Glickman, Ross, and Ross-Gordon (2007), “we cannot improve education by simply legislating higher standards and higher stakes” (p. 30). Effective classrooms engage all students for enhanced learning and on-task behavior (Forbes & Davis, 2009). Instruction and learning in the classroom must be geared toward inquiry-based instruction, effective questioning techniques, and access to technology for monitored instructional remediation for struggling learners and lateral exploration for gifted learners (Edelson, Gordin, & Pea, 1999). Literature suggests the teacher can radically improve the level of student achievement: accounting for one-third of in-school factors that contribute to student learning (Rothstein, 2010). The following empirical articles intersect instructional practices with student achievement.

Lee, Linn, Varma, and Liu (2010) analyzed inquiry instruction and its impact on classroom outcomes. The researchers set out to expand the literature by analyzing “the robustness of inquiry instruction across science topics and across teaching contexts:” six science courses and 27 teaching contexts (p. 71). In the end, the researchers sought to answer the following research questions: (a) “What is the impact of typical versus inquiry instruction on student knowledge integration across science courses and teaching contexts?” and (b) “How do teaching contexts impact student progress in knowledge integration?” (p. 72).

In framing the study, Lee et al. (2010) defined inquiry as the “learning experiences

that engage students in various combinations of identifying questions, collecting and interpreting evidence, formulating explanations, and communicating their findings” (Duschl, Schweingruber, & Shouse, as cited in Lee et al., 2010, p. 71). The researchers utilized the knowledge integration framework to implement the design of the study’s ten inquiry science units (Bransford, Brown, & Cocking, as cited in Lee et al., 2010, p. 72). The researchers predicated their study on the knowledge integration framework: the notion of students cementing knowledge through a multidimensional perspective. Using targeted professional development, the researchers, following year one of typical teaching, tailored the professional development for each teacher to elucidate the *Knowledge Integration Design Principles*.

In the study, the researchers used a delayed cohort comparison design study, whereby Lee et al. (2010) analyzed one cohort of student scores after a year of “typical” science instruction versus another cohort’s scores after a year of “online, visualization rich inquiry units” (p. 71). After the first year of typical science instruction, the researchers guided teachers through a job-embedded professional development model. Lee et al. ascertained the estimation of student knowledge from a cohort at the end of the first year of typical instruction ($n = 2,060$). Utilizing the same newly trained teachers, the researchers gathered student knowledge from a cohort at the end of the next year of inquiry instruction ($n = 2,685$).

The participants included 27 teachers from 3 middle schools and 7 high schools in 3 states. Teachers participating were 52% female and 52% teaching in middle schools. Ranging from teaching experience between 1 and 35 years, the mean teaching tenure equaled 12.9 years. Compared with other schools in the state, based on state-administered tests, three schools were classified as high achieving, five schools as average, and two

schools as low (Lee et al., 2010, p. 79). One-third of the schools serviced a high percentage of struggling science students. Ten inquiry units were developed by 11 of the 27 teachers after a 5-day professional development workshop, following the typical cohort year.

Lee et al. (2010, p. 79) determined students' knowledge integration estimates through the application of an Item Response Theory analysis: based on a Rasch Partial Credit Model; ability estimates were further ascertained through a *mean/sigma* method. Following both teaching unit years, the researchers administered the teacher surveys. The survey utilized 18 items for both inquiry and typical teaching practices. Teachers responded on a 3-level Likert-type Scale (*very important, somewhat important, not at all important*). Following the round of surveys, 12 out of the 27 teachers were interviewed concerning "general teaching and assessment strategies, inquiry unit implementation, and beliefs about inquiry-based teaching" (p. 80). Finally, the Inquiry Cohort responded to items concerning demographic and classroom instruction methods for integration of student experience into teaching context.

Furthermore, the researchers (Lee et al., 2010) employed a mixed effects analysis of variance to ascertain any changes due to inquiry instruction, teaching context, and science course. For the analysis, the dependent variable was "equated knowledge integration estimates from the year-end assessments" and the independent variables included "Cohort (Typical vs. Inquiry), Course (6 science courses), and Context (27 teachers each in a different teaching context)" (p. 81). Triangulating these findings with qualitative research (teacher surveys, interview transcripts, and project records), the researchers searched for factors relating to effective inquiry instruction.

In relationship to the overall impact of instruction, the student average knowledge

integration level for the Inquiry Cohort ($M = 0.28$, $SD = 1.23$) outperformed the Typical Cohort ($M = -0.18$, $SD = 1.52$), $F(1, 24.08) = 16.66$, $p < 0.001$. From the findings, the researchers concluded inquiry learning units delivered significantly higher student outcomes than typical teaching. Further findings gathered by the researchers showed instructors with more inquiry teaching development had higher student achievement than their less-knowledgeable inquiry teaching counterparts. The researchers urged more professional development on well-designed inquiry science units, horizontally and vertically aligning curriculum, and the knowledge inquiry framework as a means of teaching pedagogy.

A limitation for this study includes the following suggestion by the authors (Lee et al., 2010): further testing of the inquiry based units is needed to replicate findings over a wider range of classroom contexts. From this research, one infers the integral importance of inquiry-based, hands-on, technology-enhanced learning for greater student outcomes on any measure of assessment. Perhaps most important, highly trained, professionally developed teachers must exist in order to enact such high achievement in the classroom.

Aguiar, Mortimer, and Scott (2010) analyzed the classroom interactions of students precipitated by students' "wonderment questions" (p. 174). The research was predicated upon these questions' potential to encourage intellectual discourse and, subsequently, classroom engagement. The researchers investigated the following question in their research: "How do student questions impact upon the teaching explanatory structure and modify the form of the ongoing classroom discourse, in selected science lessons?" (p. 174). The purpose of the study was to analyze teaching that allowed students to ask probing questions, and its subsequent impact on class discourse (p. 176).

In grounding the research, Aguiar et al. (2010) identified two major modes of

questions: basic information questions and wonderment questions. Contained within the basic information questions, two sub-types existed: factual and procedural. At the other end of the spectrum, Aguiar et al. (p. 175) defined *wonderment*: wonderment questions elicited higher order thinking skills: “predictions, explanations, and causes...”. Aguiar et al. (p. 175) stratified wonderment questions into five levels: (a) comprehension questions for the unknown, (b) prediction questions for comprehension unknown questions, (c) anomaly detection questions for something not understood, (d) “why-is-this-important-to-me” questions, and (e) “what-should-I-do-next” questions.

In their analytical framework, Mortimer and Scott noted the impact of students’ questions in three ways: teaching purposes, patterns of interactions, and communicative approach (as cited in Aguiar et al., 2010). The research methodology included the qualitative case studies of three classroom episodes extracted from two separate teaching sequences. The study was unusual insofar as its participant also served as researcher. The two classrooms (one seventh grade class and two ninth grade classes) analyzed purposefully were selected for the frequency of student wonderment questions asked. The researchers also conducted qualitative research through the ethnographic methods of video recording and subsequent transcription and coding.

The researchers (Aguiar et al., 2010) found the authoritative or dialogic nature of the classroom discourse did not solely rest in the hands of the teacher, eschewing the “sage on the stage” mentality. Rather, the authors perceived the teacher as more of a gatekeeper to knowledge, vacillating between the dialogic and authoritative, as the teacher felt necessary in a given lesson. The researchers also alluded that teaching self-efficacy increased once teachers understood more about the types of questions that elicited a dialogic atmosphere: “‘take the risk’ of encouraging wonderment questions from their

students” (p. 190).

The focus of this study included extrapolating the types of questions teachers and students ask in class and the subsequent student outcomes, furthering the effort of a dialogic atmosphere in the classroom, and affording teachers to redirect explanations to questions within this dialogic context. From this study, one observes the high premium of an “open” classroom and its subsequent effects on student self-efficacy; therefore, professional development for teachers presents a logical solution for dialogic training.

In comparing the purpose of the two empirical studies, Lee et al. (2010) analyzed inquiry instruction and its impact on classroom outcomes. The researchers looked to expand the literature on the impact of inquiry instruction in science and other academic areas. In the research of Aguiar et al. (2010), the researchers similarly looked to investigate the impact of science pedagogy (e.g., probing questions) on enhanced student outcomes.

The content of both the Lee et al. (2010) empirical study and the Aguiar et al. (2010) research focused on inquiry instruction and how it influences student learning of concepts. However, where Lee et al. (2010) integrated science with other academic areas when investigating questions during classroom discourse, Aguiar et al. (2010) focused specifically on science when examining the effects of inquiry instruction.

In terms of methodological procedures, Lee et al. (2010) used a delayed cohort comparison design study where a cohort of teachers was measured by student achievement after no experimental training and then compared the student achievement of the following year, after the same teachers received tailored professional development, grounded in the *Knowledge Integration Design Principles*. On the other hand, the Aguiar et al. (2010) research was a qualitative case study of three classroom episodes during two separate

teaching terms. The Aguiar et al. research was unique in using the researcher as the participant.

Regarding the results of the two studies, Lee et al. (2010) found an overall advantage of inquiry instruction over typical methods when teaching science units. In terms of knowledge integration of the two cohorts, the Inquiry Cohort ($M = 0.28, SD = 1.23$) was significantly higher than the Typical Cohort ($M = -0.18, SD = 1.52$), $F(1, 24.08) = 16.66, p < 0.001$. Lee et al. emphasized that the impact of inquiry instruction, per the results of the Inquiry Cohort, should be particularly noted since the inquiry instruction lasted five class periods and was measured at the end of the year, not immediately after instruction (p. 81). In examining the results of the Aguiar et al. (2010) study, the researchers suggested the current instrument focused heavily on the role of the teacher, without analyzing interactions spearheaded by student questions. Also, Aguiar et al. discussed the complexity in the tension between the teacher authoritative role when introducing the science topic and the teacher dialogic role of instruction where the teacher allows questions and curiosity to drive student learning and dispel scientific myths and misconceptions.

In analyzing the limitations of the two empirical studies, the Lee et al. (2010) research could have been marked with validity issues through the use of an inconsistent sample. Lee et al. used one group of students as the sample ($n = 2,060$) for a cohort during the first school year, with teachers using typical instruction, and an entirely different set of students ($n = 2,685$) during the second school year, with teachers using inquiry instruction. While the study did span three states and 27 different teachers in both middle and high schools, the use of roughly 2,000 different students from year one to year two could pose a distinct threat to validity. In terms of the Aguiar et al. (2010) study, the researchers

employed a small sample size (one 7th grade class and two 9th grade classes) with only one teacher working with all three classes. Both the small sample size and teacher participants used in the research of Aguiar et al. could have posed threats to validity and subsequent generalizability. As noted above in the paragraph on results, the generalizability of the Aguiar et al. research fails to extend to the more “personalized learning” type classrooms where student questions and curiosity drive instruction—with questions in this study representing an extraneous variable that might obfuscate clear causality of this research, which would in turn be a threat to validity.

This author gleaned a sense for direction of the field in both the articles of Lee et al. (2010) and Aguiar et al. (2010) in that both call for extensive professional development for teacher training in the areas of facilitating inquiry-based learning and dialogue in the classroom. Perhaps, according to Aguiar et al. (2010), more emphasis should be placed on professional development that allows teacher to practice and hone “opening up” to specific science topics, to develop inventories of common scientific questions of students, and effective ways to respond. Both researchers encouraged further empirical study in the area of inquiry based instruction, with Lee et al. (2010) encouraging research with other frameworks and with more participants and Aguiar et al. (2010) calling for examination of the participants’ role in the discourse. In the following sections, the researcher investigates additional specific trends in pedagogy for the elementary science classroom.

Nature of science. Morrison, Raab, and Ingram (2009) investigated how elementary teachers differed from secondary teachers in their understanding of the Nature of Science (NOS). Abd-El-Khalick and Lederman noted that teachers who teach science often lack the proper background knowledge about the NOS (as cited in Morrison et al., 2009, p. 384). Many teachers had not experienced scientific research or even known

scientists. A challenge existed for professional development coordinators to design science training to improve teachers' views of the NOS.

The purpose of the current study was to decide how elementary and secondary teachers' opinions of NOS and scientists were impacted through a professional development program that immersed the teachers in a research environment surrounded by examples of the NOS. The objective was to have teachers thinking about science as a process, discussing the characteristics of the NOS with scientists, and identifying samples of these characteristics at the research center. The researchers designed the professional development for teachers to talk with scientists on a variety of topics and view what scientists do on a daily basis. Morrison et al. (2009) developed the following research questions: (a) How did elementary and secondary teachers' views about the NOS change from the beginning to the end of the professional development experiment? (b) How do daily interactions focused on the NOS (discussions, interviews, job shadowing, etc.) impact elementary and secondary teachers' views of scientists? and (c) How do daily interactions focused on the NOS (discussions, interviews, job shadowing, etc.) impact elementary and secondary teachers' ideas regarding teaching science? (p. 387).

Contextually speaking, the professional development took place at the Laser Interferometer Gravitational-Wave Observatory (LIGO). The professional development module focused on two sessions in a joint collaborative graduate course, offered for science teachers, by a midsized, western university and the LIGO Observatory (Morrison et al., 2009). During a two-week session of the graduate course, science teachers spent time shadowing a scientist. During this time frame, the following characteristics of science were emphasized: (a) the changing, yet reliable nature of science, (b) the empirical nature of science, (c) the subjectivity of science, (d) the differences between scientific theories

and laws, (e) the human influence on science, (f) the social and cultural aspects of science, and (g) the observational and inference-laden qualities of science (p. 387). During this same time frame, teachers actively engaged in discourse with the scientists, sharing their interview notes and conversations with fellow colleagues. Also, during the two-week collaboration with the scientists, the teachers participated in inquiry-based learning activities. Finally, their summative assessment included the following: (a) presentation of one inquiry activity, (b) four inquiry activity papers, (c) an inquiry unit plan, and (d) a summative paper concerning how their perspectives had ultimately changed as a result of their collaborative activities (p. 388).

Since the study took place over two summers, with different participants (but with same content presentations), the researchers cited the total number of subjects as 20 ($n = 20$). During the study, the participants were asked if they would actively participate, be willing to complete questionnaires, and be interviewed one to two months following the course. Data were collected in a variety of methods during the 2 weeks of each summer course: pre- and post-NOS questionnaires, interviews, written work, class discussions, presentation of the science activity, and researcher's notes. The researchers used the *Views of Nature of Science Questionnaire-Form B* (VNOS-B) by Lederman, Abd-El-Khalick, Bell, and Schwartz (as cited in Morrison et al., 2009, pp. 388-389) for quantitative data.

In analyzing the data, Morrison et al. (2009) created a profile of each participant based on the aforementioned data tools. Based on the data profile, each participant was ranked according to their views of the seven postulates of the NOS as *an informed view*, *an adequate view*, and *an inadequate view*. In regards to Research Question 1, pre- and post-questionnaire data exemplified an enhanced knowledge of NOS. In regards to Research Question 2, the teachers' views about scientists were not significantly changed through the

collaboration; however, through the collaboration, the teachers' own insights into their teaching methodologies did change. According to Morrison et al., the teachers began to observe science as their own students observed science, following their collaboration with the scientists.

Through this study, Morrison et al. (2009) remarked that teachers now had real-world examples to draw upon in instructing their own students: making the content come to life. Also, from the interview process with scientists, teachers began to discover what kept those scientists interested in the classroom during their formative years—yet another tool for engaging students in the classroom.

From this observation, implications for service teachers include allowing the students, either first-hand or virtually, to experience collaborative efforts and discourse with scientists in the workplace—answering the students' question “why do I need to know this?” For teachers and even students, Morrison et al. (2009) commented that true professional development and teaching, respectively, need not involve empirical research, but moreover, the authentic opportunity to observe and engage in an environment where science surrounds them, delivering authenticity to the research environment (p. 400).

Akerson, Cullen, and Hanson (2009) researched the relationship between a Community of Practice (CoP) professional development program and teacher views of the Nature of Science (NOS). The researchers documented a qualitative case study of three teachers to depict changes in views and teaching practice. The processes and beliefs (NOS) by which scientists work and approach problems have been outlined to consist of several components. The research question examined how a CoP professional development program influences (a) teachers' views of NOS and (b) NOS teaching practice.

The researchers based their theoretical perspective on the work of Putnam and Borko (as cited in Akerson et al., 2009, pp. 1091-1092). According to Putman and Borko, cognition and learning are found in physical and social contexts, are social in nature, and are distributed among the individual, other persons, and tools (as cited in Akerson et al., 2009, pp. 1091-1092).

Akerson et al. (2009) used the NSTA Position Statement for K-12 students. The NOS components included the following: (a) scientific knowledge was both reliable and tentative, (b) no single method existed, but there were shared characteristics of scientific approaches, (c) creativity played a role in the development of scientific knowledge, (d) there was a crucial distinction between observations and inferences, (e) though science strives for objectivity, there was always an element of subjectivity (theory ladenness) in the development of scientific knowledge, and (f) social and cultural contexts played a role in the development of scientific knowledge (NSTA; as cited in Akerson et al., 2009, p. 1092).

The study included a 12-month professional development program created to form a CoP for 17 K-6 grade teachers (Akerson et al., 2009, p. 1093). The participants were female Caucasians who varied in years of teaching experience and science teaching ability. The researchers chose three school districts located near the university to partner in the study. The district demographics included high free and reduced lunch populations and served both rural and suburban populations. Nine different elementary schools, from the three districts, participated in the professional development program.

The teachers met regularly to develop and share common values and goals and participate in collaboration and reflection of the work. The researchers focused on teacher and student learning, emphasized goals, encouraged conversations on teaching and learning issues, and suggested that no “perfect” teaching method exists. Teachers attended

state and national conferences and shared the learning with teachers at the home schools. The program extended into the summer to introduce NOS ideas and guided inquiry. The teachers developed physical science units to use with students (Akerson et al., 2009).

The researchers rotated the school year training at multiple school sites. Teachers planned and presented the work at science teacher conferences to increase the global nature of community membership. The teachers spent time discussing the challenges and benefits of teaching NOS. By sharing, teachers actively shared in the CoP process. Akerson et al. (2009) modeled inquiry activities to help teachers resolve problems. The two classroom visits allowed the researchers to observe the various ways that teachers focused on NOS.

The study examined the relationship between elementary teachers' views of NOS and teaching practice. The authors used an interpretive model to determine the influences (Bogdan & Biklen, as cited in Akerson et al., 2009, p. 1101) with a variety of data sources that informed the ongoing professional development. Pre- and post-summer workshop and post-school year assessments using the Views of Nature of Science Questionnaire-Form VNOS-D2 (Lederman & Khishfe, as cited in Akerson et al., p. 1101) examined the changes in NOS views attributed to the CoP participation. The researchers interviewed 30% of the randomly selected participants prior to the program, at the conclusion of the summer workshop, and at the end of the school year. The transcribed audiotape interviews provided interviewees the opportunity to expand on views and the researchers to validate the interpretation of teachers' written responses. Open-ended questions during the interview addressed the teachers' perceptions of the program, teaching practices, classroom observations, and insights into improving the program. The researchers collected field notes during each classroom visit to document the emphasis of NOS within science teaching. Teachers provided lesson plans and student work to assist in the NOS emphasis.

During the summer workshop, the researchers collected videotapes of teacher interaction (Akerson et al., 2009, pp. 1101-1102).

The researchers independently analyzed NOS interview transcripts and questionnaires to validate the interpretation of teacher views. Coders documented common misconceptions of the teachers' views including stating that the "scientific method is the only way to create scientific knowledge" (Akerson et al., 2009, p. 1101). The researchers categorized the NOS views and checked for confirmatory or contradictory evidence in the data. The researchers conducted multiple rounds of "coding, confirmation, and modification" to reduce the data (p. 1101).

Akerson et al. (2009) reviewed the field notes and videotapes to determine the nature of interactions among teachers and among teachers and researchers. The researchers looked for indicators of the development of a CoP by evidence of mutual engagement. Akerson et al. triangulated how teachers incorporated NOS and discussed the information with other teachers. The researchers also examined the evidence of whether the teachers emphasized the NOS views and how the instruction changed over time.

The results showed all teachers' views of NOS improved. Teachers taught NOS in a variety of ways indicating that the relationship between knowledge and practice is more than just linear. Some teachers used NOS to emphasize science teaching while others formulated NOS strategies rather than using it as an overarching theme. The teachers noticed changes in NOS views and discussed the changes in the professional development community. The CoP developed a well-supported community of learners and assisted teacher change when combined with the NOS modeling and reflection.

The researchers discussed the importance of providing extended time to assist teachers in improving their NOS views. Encouragement is necessary for teachers who tend

to operate in silos and additional support is paramount for the teachers who struggle. Through the Community of Practice professional development model, teachers were afforded a springboard for developing a NOS consistent with accurate knowledge and information. In essence, in order to effect change for a deeper scientific understanding in students, teachers must also be prepared to conduct ongoing and job-embedded professional learning. In the following section, the researcher discusses another pedagogical specific methodology for instruction in the elementary science classroom.

Inquiry-based learning. In examining inquiry-based learning, Bybee (2002) discussed connections between how students learn and scientific inquiry, identifying three main relationships: (a) learning is a basic, adaptive function of humans; (b) learning originates in diverse experiences; and (c) learning is a process that can be enhanced through teacher facilitation (p. 28). In terms of the first association, very young children begin making sense of the world as a direct result of learning processes. Bybee continued with the assumption that young students even possess an inclination to learn and conceptualize ideas regarding “biological and physical causality, number, space, time, and language” (p. 28).

In regards to science learning originating through diverse experiences, Bybee (2002) dichotomized the science knowledge acquired through everyday, organic inquiry by students versus the traditional classroom (p. 28). While student explanations of phenomena were incomplete, when compared to the learning later taking place in the science classroom, Bybee noted the importance of students bringing in preconceptions to the classroom—noting that students hold on to misconceptions about science, even when confronted with the accurate, contradictory information from the classroom (Bybee, 2002, p. 28). Thus, Bybee suggested that it is only through the integral interactions of students,

peers, parents, and friends—oft facilitated by the teacher—that students can truly reconcile preconceptions and misconceptions with accurate information, as students question what is personally ascertained outside the classroom with what is learned inside the classroom.

For the final linkage between learning and scientific inquiry, Bybee (2002) cited several different pedagogical methods for teachers to facilitate scientific learning: (a) identifying current conceptions of students; (b) challenging the adequacy of current conceptions by students; (c) introducing concepts that are understandable and accurate; and (d) affording a wealth of opportunities to apply new concepts to familiar student constructs (p. 28). Bybee noted that any learning of new scientific concepts required ample amounts of time and opportunities to construct the new learning. Ultimately, Bybee declared that students learning science required “content that is deeper than facts and information, a curriculum that is richer than reading, instruction that is longer than a lesson, and teaching that is more than telling” (p. 29).

Geier et al. (2008) investigated the reform movement in urban middle schools that focused on standard based science teaching. The specific context took place in the Detroit Public Schools and was centered on project-based, inquiry learning science units, bolstered by specific professional learning opportunities and learning technologies.

According to the researchers’ framework, inquiry learning challenged teachers to “develop new content knowledge, pedagogical techniques, approaches to assessment, and classroom management” (Geier et al., 2008, p. 922; cf, also, Edelson, Gordin, & Pea, 1999). Lee and Luykx (2004) remarked on a deficit in the literature on administering inquiry-based units to diverse populations for this strand of research (as cited in Geier et al., 2008, p. 922). In light of these problems, Geier et al. (2008) posed the following core question: “Whether [or not] urban student participation in project-based inquiry science

curricula leads to demonstrably higher student achievement on statewide assessments over and above general district-wide efforts at reform?” (p. 925).

Demographic factors such as a professional staff of 10,000, with a student population of 160,000 (91% African-American students, 5% Latino, and 4% “other”) describe the population. Sixty-nine percent of students were eligible for free and reduced lunch. During the three years of implementation, ending with the 2000-2001 school year, 37 teachers, and approximately 5,000 students, participated in the study. Selection criteria for schools included “adequate technology, supportive administration, and equity among schools in access to innovative programs” (Geier et al., 2008, p. 927). Each year of the study, the inquiry-based methods were scaled up within the Detroit Public Schools, adding more schools and teachers to the participant pool.

The Michigan Educational Assessment Program (MEAP) was the high-stakes, state-administered test utilized in the study for science achievement. The Center for Learning Technologies in Urban Schools (LeTUS) spearheaded the specific inquiry-based units taught by the designated teachers. The method of analysis was a casual comparative design between those students participating in the LeTUS units and those who did not. Subsequently, the researchers compared the treatment sample students in Cohort I ($n = 760$) and Cohort II ($n = 1,043$) to comparison groups of 8,900 and 8,662, respectively.

In analyzing the data, both cohorts exhibited significantly higher test scores than the comparison groups in both science content understanding and process skills. Gains were measured a year and a half after participation in the inquiry-based learning units. The high scores did not diminish for the second cohort, even when scaling up took place and the number of teacher participants increased. The researchers concluded the positive growth of students in each cohort group was significant, $p < 0.001$. Participation in one

cohort's LeTUS unit equaled a significant 14% improvement in total score on the MEAP, when compared to the comparison group. Participation in at least one LeTUS unit increased a student's passing rate on the MEAP by 19% in Cohort I and 14% in Cohort II (Geier et al., 2008).

The Detroit Public School district-wide MEAP scores revealed statistically significant differences in gender, with boys scoring 17 points below girls in 2000 ($t = 7.45$, $df = 9,544$, $p < 0.001$, $ES = 0.15$) and 11 points below girls in 2001 ($t = 5.01$, $df = 9,631$, $p < 0.001$, $ES = 0.10$). The aforementioned data held true in light of research that expostulated boys underperform girls in predominantly African-American populations (Graham; Lynch; Weaver-Hightower; Witherspoon, Speight, & Thomas; all cited in Geier et al., 2008, p. 932). The researchers however concluded that participation in at least one LeTUS unit narrowed the achievement gap of genders. Geier et al. found the following: "standards-based instruction including pervasive technology and project-based units appears to engage at-risk urban male learners, narrowing and closing the gender gap in achievement with their female peers" (p. 933).

Geier et al. (2008) found that comparable gains were achieved across different grade levels, with high participation equaling higher achievement outcomes. Further positive findings of the study included a narrowing of the gender achievement gap exhibited by African-American boys. The researchers noted the positive and integral impact of inquiry-based learning units on the achievement scores of students. Even in light of "historically underserved urban students," when "the curriculum is highly specified, developed, and aligned with professional development and administrative support," students succeeded—not only increasing achievement gains, but, also, intrinsic motivation (p. 922).

One limitation of the study was potential sample bias: “potential influence of student absenteeism and attrition, biases in student selection during tracking, and possible school site and participant teacher selection factors” (Geier et al., 2008, p. 928). The low numbers of the treatment sample could have served as another limitation. Geier et al. agreed that the students who did remain in the LeTUS units could possibly be more academically able. Implications include a focus on job-embedded professional development utilizing inquiry-based units for narrowing the achievement gap between males and females.

Wilson, Taylor, Kowalski, and Carlson (2010) performed a laboratory-based randomized control study to examine inquiry-based pedagogy in the science classroom. The researchers wanted to ascertain whether commonplace instruction and inquiry-based instruction afforded equal learning opportunities. The research problem was the inconclusive nature of evidence for inquiry-based instruction and resources, in light of the federally mandated evidence-based reform initiatives.

Wilson et al. (2010) designed the study to test the differences in achievement between students who received inquiry-based instruction using the BSCS 5E Instructional Model and students who received same-content instruction through commonplace teaching practices, as designated by national surveys (p. 279). The BSCS 5E provided an instructional model in teaching and leading student inquiry. The researchers measured the three following student outcomes: (a) scientific knowledge, (b) scientific reasoning through application of models, and (c) construction and critique of scientific explanations (p. 281).

Wilson et al. (2010) listed the following research questions: What is the effect of inquiry-based materials on student achievement as compared to commonplace outcomes?

This question was sub-divided into the following: (a) To what extent can differences in student achievement between the inquiry-based and commonplace groups be attributed to randomized group assignment? (b) Does student race/ethnicity, gender, or socioeconomic status account for variation in posttest scores above and beyond variation accounted for by pretest scores and group assignment? and (c) What differences in achievement by treatment group exist specific to the learning goals of knowledge, reasoning, and argumentation? (p. 281).

In the methodology section, Wilson et al. (2010) listed a laboratory-based randomized control study design. The sample was selected from an invitation to Colorado Springs area schools, youth organizations, and home-school groups for children between the ages of 14 and 16 to participate in research over a two-week time period with a 14-hour commitment. Fifty-eight students were selected and a coin-flip determined their placement in either the Inquiry-Based Unit ($n = 30$) or Commonplace Unit ($n = 28$). Both units were taught by the same teacher in a controlled setting in the BSCS classroom in Colorado Springs, Colorado.

The commonplace unit instruction included the following facets: (a) an emphasis given to various instructional objectives, such as learning terms and facts, learning to evaluate scientific arguments, or learning about the nature of science; (b) frequency of teachers' use of various instructional strategies, such as introducing content through formal presentations, posing open-ended questions, or asking students to consider alternative explanations; and (c) frequency of student participation in various activities, such as watching a demonstration, following specific instructions in an activity or investigation, or designing and implementing their own investigation (Wilson et al., 2010, p. 282).

With only several modifications to the inquiry-based unit, the authors listed the

following hallmarks of the BSCS 5E model: students (a) ask questions such as, “Why did this happen?” “What do I already know about this?” “What can I find out about this?” and show interest in topic; (b) test predictions and hypotheses, record observations and explanations, and form new predictions and hypotheses; (c) use recorded observations in explanations and develop explanations based on data; (d) form new predictions and hypotheses, try alternatives and discuss them with others, compare personal explanation with scientifically accepted explanation, and assess own understanding; and (e) explain possible solutions or answers to others, listen critically to others’ explanations, question others’ explanations, and check for understanding among peers (Wilson et al., 2010, p. 280).

Data were collected from identical pre- and post-tests administered immediately before and after instruction and a thirty-minute interview four weeks after the instructional unit. The tests contained multiple-choice, true/false, and constructed response items: the multiple-choice and true/false questions had a mean item difficulty of 0.789 with a Cronbach’s alpha of 0.695 for the total test reliability index. The thirty-minute interview was formed around open-ended questions that covered the concepts taught in the units. Both instructional groups were observed by three researchers who took anecdotal records and filled out an observation protocol. Upon completion of the unit, all students took a 17 item survey. Each class session was also filmed and coded (Wilson et al., 2010).

The observation protocol scores for the inquiry-based unit were higher than those of the commonplace unit across [$t(48) = 9.937, p < 0.01$] and within protocol subscales ($p < 0.01$ for each subscale). Also, the completed survey mean scores for the inquiry-based unit were higher than the commonplace unit [$t(55) = 3.195, p < 0.01$]. In regards to the classroom recordings, total minutes tallied showed time spent on lecture was higher for the

commonplace instructed unit. In contrast, higher levels of critical thinking were observed in the inquiry-based instructed unit. Wilson et al. (2010) noted a key difference in that the inquiry-based group was *constructing knowledge* through writing activities, while the commonplace group was only *receiving knowledge* (p. 286). Four raters scored a sample of 10% of the total tests using the intraclass correlation coefficient to determine inter-rater reliability. No dramatic difference was detected among raters [$F(1, 47) = 0.033, p = 0.992$] with an intraclass correlation coefficient of 0.783 (two-way mixed effects model, single measure, absolute agreement).

In relationship to total test scores, the inquiry-based group exhibited much higher posttest scores than the commonplace group [$F(1, 55) = 4.570, p < 0.05$]. The effect size for the difference was 0.47 standard deviation units (Cohen's d). In analyzing the constructed response items on the tests, Wilson et al. (2010) found the inquiry-based group had significantly higher scores than the commonplace group, [$F(1, 56) = 4.537, p < 0.05$]. The Cohen's d effect size for the difference was 0.68. Free and reduced lunch status, race/ethnicity, and gender did not contribute to much variation in posttest scores when weighed against other factors. When analyzing the argumentation scores for the students, the researchers found much higher scores (from the interviews) in the inquiry group for claims [$F(1, 54) = 4.253, p < 0.05$], evidence [$F(1, 54) = 9.794, p < 0.01$], and reasoning [$F(1, 54) = 5.051, p < 0.05$]: with Cohen's d effect sizes of 0.58, 0.74, and 0.59, respectively. In the discussion section, the researchers noted students in the inquiry-based group attained higher levels of academic achievement than their commonplace instructed counterparts. This observation was detected across the learning goals of knowledge, reasoning, and argumentation. Also, inquiry-based instruction proved more effective in both time frames: immediately following instruction and four weeks later. Commonplace

science instruction led to academic outcome gaps by race. Inquiry-based instruction did not show these same racial gaps.

One possible limitation was that the “BSCS” classroom was utilized for instruction in both the inquiry-based and commonplace group. Thus, the arrangement of the room and materials could have affected the control group, possibly skewing the results.

Furthermore, since the unit was on sleep and sleep cycles, no real hands-on activities were conducted in the inquiry-based group, a hallmark of that instructional style. Wilson et al. (2010) indicated that the research lends itself to promoting more inquiry-based instruction in the science classroom; the value of quality, inquiry-based curriculum in science instruction; minimizing achievement gaps by race through inquiry-based learning; and enhancing academic outcomes in the age of reform and accountability. This study presents strong evidence for the implementation of inquiry-based instruction in the science classroom—evidence that helps meet the growing need for hard data to guide instruction for developing lifelong learners in STEM subjects.

Palmer (2009) investigated “situational interest” and its sources as the primary purpose of his study. Rennie, Goodrum, and Hackling (as cited in Palmer, 2009, p. 147) proffered that many students lacked motivation in the classroom setting, especially in science which they considered boring. Palmer deduced that a combination of unmotivated pupils and student perception of science as dull diminished the learning outcomes in the science classroom.

Palmer (2009) focused his study on situational interest, defined as a fleeting moment (opportunity) for learning; yet, with multiple experiences of situational interest, long-term interest could occur. Thus, Palmer’s research problem examined situational interest in science classrooms. The two research questions were the following: (a) how

much situational interest is generated during different parts of a science lesson? and (b) what are the sources of situational interest?

The sample was 224 ninth grade students ($n = 224$) of ages 14 and 15 who went to five different schools in a south-eastern Australian city. The sample was selected in the following manner: science classes were ability grouped and the same number of student volunteers was taken from every ability group of every ninth grade science class, pulled from the five school sample. The study used a purposefully selection of eight students from every class, with gender demographics relatively equal and with a range of socioeconomic backgrounds (Palmer, 2009).

Each eight student group was taught the same 40-minute lesson during 28 class periods by the same instructor to diminish bias. During each lesson, the teacher first demonstrated a hands-on activity and solicited comments from the class on the specific content being taught. The teacher then proposed a variety of other resources available for an experiment on the content. The teacher next allowed students to select their materials, complete the experiment, and make observations. As student groups finished experimenting and recording observations, each pair of students gave oral reports to share findings with classmates—since each group could possibly be investigating different phenomena within the content area. Palmer (2009) developed a baseline level of interest by making the students copy notes at the beginning and end of each lesson.

The design study was a hybrid qualitative/quantitative design. To streamline data collection and analysis, the researcher used only one Likert-type item to measure interest level: “I thought this part was...” (1 = *very boring* to 5 = *very interesting*). Students responded to this item after first copying notes, demonstration, proposal, experiment, report, and the final copying of notes. Data were analyzed through a one-way ANOVA.

Two-way ANOVAs were employed to determine effects of gender and achievement level. Qualitative data were collected from group interviews conducted at the end of each lesson. Much like the quantitative item, Palmer (2009) asked students to comment on their level of interest at each of the six levels of the lesson. For reliability, a representative sample of forty-seven responses were coded, with agreement found in 87% of cases. The validity of the one-item instrument was confirmed through the comparison of the students' interest levels reported on the written responses to the oral responses from the group interviews. Reliability was obtained through comparing the scores from the two sets of note-taking phases. Across-group reliability was deduced from using multi-site data gathering.

In analyzing the data, the researcher noted that the hands-on inquiry skills of the students were at a low level, possibly due to lack of experience in investigation. The results showed interest arousal was a significant result of inquiry skills lessons—with fluctuation occurring depending on the types of activities involved. Novelty produced the highest arousal in interest. Choice, physical activity, and social involvement were also implicated. In an analysis of overall interest levels, the one-way ANOVA gave an F -value of 315.57, with $p < 0.05$. In terms of achievement, Palmer (2009) concluded that higher achieving students were only slightly more interested in the lessons than their lesser achieving counterparts.

In disaggregating the qualitative data from the audio-recorded interviews, “learning” was the most common source of interest: 71% reporting interest in the demonstration phase, 79% reporting interest in the report phase, and 39% reporting interest in the experiment phase. Students also noted “choice” as a source of interest, with 68% reporting interest in the proposal phase. Finally, “novelty/suspense/surprise” was a large source of interest for students, with 57% reporting interest in the demonstration phase and

46% reporting interest in the experiment phase. However, in the experiment phase, “physical activity” was the most common source of interest, with 54% reporting interest. The researcher concluded that students were most interested when engaged in inquiry skills such as proposing investigable questions, making observations and explanations, and reporting (Palmer, 2009, p. 159). Even in light of student lack of experience with asking investigable questions, observing, and explaining, scientific inquiry tasks did indeed spark situational interest.

Limitations for this study (Palmer, 2009) include the small sample size, eight students as opposed to full classes. Also, longer amounts of time are usually allotted for inquiry lessons, while the researched lessons from the study only lasted forty minutes in duration. Implications for learning include the motivational methods inherent in sparking situational interest through inquiry learning. While agreeing that motivation is not enough, Palmer (2009) affirmed that the amalgamation of motivational strategies and sound instructional practices resulted in enhanced student outcomes in the science classroom. According to this study, engaging students in content and research-based instructional pedagogy through the use of inquiry lessons improves assessment results in this age of accountability. Furthermore, at the core, sparking situational interest speaks volumes about the type of learning this nation wants to promote early in life.

Throughout the current section on Elementary Science, the researcher noted best practices in science that aligned with the three SISI Academic Performance standards: Standard 1, Curriculum; Standard 2, Classroom Evaluation/Assessment; and Standard 3, Instruction. By embedding best practices in science at the elementary school level, classrooms, and subsequently students, are afforded a wealth of research-based pedagogy that drastically enhances student achievement. In the following section, the researcher

examines the nexus of Academic Performance, Accountability, and Science Outcomes, consequently framing the relevant deficits in the research on these topics.

Academic Performance, Accountability, and Science Outcomes

The intersection of Academic Performance, Accountability, and Science Outcomes represents the exact point whereby the current research stands to address the dearth of current studies in the field. More specifically, the current research examines how the three Academic Performance standards from the SISI—Standard 1, Curriculum; Standard 2, Classroom Evaluation/Assessment; and Standard 3, Instruction—impacted school accountability on elementary science score outcomes, as measured by the Kentucky Core Content Test from 2001-2005, with the assumption that better classroom academic practices enhance science achievement measures. In the following section, the researcher presents the theoretical framework through which to ground the current study.

Theoretical Framework

For the selection of a theoretical framework, the researcher chose Systems Theory to ground the current study. Laszlo and Krippner (1998) defined systems as a “complex of interacting components together with the relationships among them that permit the identification of a boundary-maintaining entity or process” (p. 2). While systems in the natural world are easier to identify and define, human and conceptual systems are more difficult to approximate since they can change over time (p. 2). Furthermore, artificial systems pose challenging for identifying overarching trends since a human system contains various and sometimes overlapping purposes: (a) the purpose of the system, (b) the purpose of its parts, and (c) the purpose of the overarching system, the suprasystem (p. 2). In essence, systems theory examines the constituent parts of a system and how those parts influence the “characteristic functions, properties, and relationships that are internal or

external to the system” (p. 2).

Capra (1997) cited systems theory as a holistic approach, an overarching lens, through which researchers examine any natural system. A characteristic feature of systems theory is the ability of researchers to use this theoretical framework to investigate phenomena across a variety of disciplines, using insights from each respective discipline (Mele, Pels, & Polese, 2010, p. 127). According to Laszlo and Krippner (1998), one major advantage of systems theory is the ability to provide a “trans-disciplinary framework for a simultaneously critical and normative exploration of the relationship between our perceptions and conceptions and the world they purport to represent” (p. 4). Furthermore, systems theory affords researchers the opportunity to comprehend “complex dynamics of human bio-psycho-socio-cultural change” (p. 4). Ultimately, systems theory should be regarded as a field of inquiry rather than specific disciplinary processes (p. 4).

Seminal systems theorist von Bertalanffy defined a system as a “complex of interacting elements” (as cited in Mele et al., 2010, p. 127) as he proposed in his “Allgemeine Systemlehre,” or general system theory (Laszlo & Krippner, 1998, p. 4). Ackoff (1981) defined a system as an entity comprised of two or more interrelated parts that possessed the following: (a) each element has an effect on the functioning of the whole; (b) each element is affected by at least one other element in the system; and (c) all possible subgroups of elements also have the first two properties (pp. 15-16). The universality of systems theory application can also be achieved through the interchange of the terms “element” and “component” (Laszlo & Krippner, 1998, p. 8). Von Bertalanffy encouraged systems thinking in all disciplines to discover overarching principles that applied to all systems (as cited in Mele et al., 2010, p. 127). Systems theory thinkers such as Checkland (1997), Weinberg (2001), and Jackson (2003) noted that moving

systematically from the part to the whole marks systems theory as an effective theoretical framework to investigate a specific phenomenon. At the heart of systems theory is the notion that to understand truly a natural phenomenon, scientists must employ a “global vision to underline its functioning” rather than only examining the constituent elements (Mele et al., 2010, p. 126). According to von Bertalanffy, systems theory affords researchers a holistic perspective: “although we can start from the analysis of the elementary components of a phenomenon, in order to fully comprehend the phenomenon in its entirety we have to observe it also from a higher level” (as cited in Mele et al., 2010, p. 126).

Mele et al. (2010) noted the focus on interactions as a hallmark of systems theory (p. 127). For the current research, with a lens for reform and the continuous improvement of achievement in an organization (e.g., classroom, school, school district), systems theory has an unmistakable impact of a foundational nature for investigating the interplay of component parts within this study. Barile and Polese discussed the viable systems approach, with its concern for viability and stability, in “creating its own internal environment that is able to respond effectively to external stimuli at all levels (*viability*);” furthermore, Barile and Polese noted the adaptation of a system through the maintenance of viability as achieved through “continual dynamic processes and several kinds of internal changes (*adaptability*)” (as cited in Mele et al., 2010, p. 131, emphasis in the original). As low-performing schools were mandated to conduct Scholastic Audits and Reviews, the internal mechanisms of reform (*Standards and Indicators for School Improvement (SISI)*, the threat of sanctions, the public perception of being categorized as under-achieving) were intended to catalyze these schools into taking action.

With a litany of moveable parts embedded in Kentucky school reform, such as the

nine standards and 88 indicators of the SISI, low-performing schools would be apt to “lose sight of the forest because of the trees.” Nevertheless, the SISI document gave educators specific insight into the general trends of best practices, in terms of the three Academic Performance standards of the current research, that helped alleviate achievement anxiety by affording a framework for school improvement. By following the roadmap outlined by the SISI document, educators, schools, and school districts could develop a contextual game plan whereby the overall system could be improved through an intentional focus on one or more the constituent components of the SISI; thus, the overall system could progress toward viability and stability.

Organizational theorist Peter Senge (2006) noted the following regarding the innate sense of inquiry:

From a very early age, we are taught to break apart problems, to fragment the world. This apparently makes complex tasks and subjects more manageable, but we pay a hidden, enormous price. We can no longer see the consequences of our actions; we lose our intrinsic sense of connection to a larger whole. When we try to “see the big picture,” we try to reassemble the fragments in our minds, to list and organize all the pieces...[and] after a while we give up trying to see the whole altogether. (p. 3)

Through the theoretical framework of systems theory, the current research is afforded an overarching construct through which to glean insight from the representative parts of the conceptual framework (see Figure 1) and, ultimately, from the reform system as a whole.

Empirical Studies

The current research expands on the set of dissertations examining the KDE

Scholastic Audit database (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) spearheaded by co-chairs Dr. Miller and Dr. Smith. In this section on empirical studies, the researcher briefly discusses each of the aforementioned dissertations.

In the first dissertation to examine the *Standards and Indicators for School Improvement* (SISI) and the Scholastic Audit process as a means of whole school reform, McKinney (2007) specifically analyzed the impact of SISI Standard 7 (Leadership), as gauged through the lens of elementary school principals, on student achievement, with Standard 1 (Curriculum) and Standard 3 (Instruction) as mediating factors. For the research, McKinney employed the following central research question: “Based on Scholastic Audits, what are the effects of Leadership, Curriculum, and Instruction on Kentucky accountability scores?” (p. 14).

McKinney (2007) used quantitative research through a secondary database of Scholastic Audit data, maintained by the Kentucky Department of Education (KDE), with the school as the unit of analysis. In terms of the significance of the study, McKinney noted several ways that his research adds to the educational literature in the field. First, the study furthered the field through research on the direct and indirect impact of school administrators on student outcomes, accounting for demographics and seen through a lens of leadership behavior (p. 18). Second, McKinney added to the field through additional information on the impact of leadership reform (p. 18). Third, with the inclusion of Curriculum and Instruction in the study, McKinney noted the following implications: professional learning opportunities, accountability systems, principal retention, and leadership development (p. 19). Fourth, the study added to the existing research base through a focus on how leadership, in an accountability environment, was mediated by curriculum and instruction. Fifth, McKinney believed his study could replace

accountability assumptions with research-based data at the elementary school level. Sixth, McKinney proffered that the study could also clarify principal success through data “designed to quantify principal leadership within the confines of Kentucky’s documented expectations for principals” (p. 20). Finally, McKinney offered that the study would help approximate the impact of principals on the quality of curriculum and instruction.

In terms of limitations of the study, McKinney (2007) suggested several ways that the generalizability of his study could be limited. With the reform measures being discussed being Kentucky-specific, McKinney believed that the generalizations could only be extended to Kentucky; however, with similar standards in other states, general leadership and accountability implications could be inferred (pp. 20-21). Also, with the evolution of the field of leadership research, McKinney noted that the study could be limited by content and methodology (p. 21). McKinney further suggested nine specific issues of limiting factors for the study: (a) only elementary school data were examined; (b) the quality of audit team trainings for the Scholastic Audits; (c) changing composition of audited schools; (d) perceptions of audit teams and performance of educators on a given Audit day; (e) lack of other educationally vested parties included in study; (f) state-specific accountability measures; (g) changing federal and state regulations; (h) focus only on academic outcomes rather than other unintended effects (enjoying learning, etc.); and (i) the lack of clear research on how administrators perceive standards-based reform measures (pp. 21-23).

For the methodology of the research, McKinney (2007) used quantitative research on a secondary database of Scholastic Audit data, maintained by KDE. The study focused only on Kentucky elementary schools ($N = 181$) that conducted a Scholastic Audit or Scholastic Review from 2001-2005 (pp. 78-79). Only elementary school configurations

that included fourth and fifth grades were included in the study, since the Academic Index at the elementary level is pulled from those grade levels. For the dependent variable, obtained from the Kentucky Performance Report (KPR), McKinney utilized the Academic Index (AI) scores from the audited elementary schools (p. 82). In terms of independent variables, McKinney employed the control variables of demographic factors, the alterable variables of Standard 7 (Leadership) from the Efficiency grouping of the SISI, and Standard 1 (Curriculum) and Standard 3 (Instruction) from the Academic Performance grouping of the SISI (pp. 84-87). McKinney used descriptive statistics for the demographic data, Scholastic Audit evidence, and Academic Index scores (p. 89). In terms of psychometric analysis, the study included factor analysis and Cronbach's (1951) alpha to confirm the validity of the set of indicators representing each standard and assure the internal consistency of these factors, respectively (pp. 89-90). McKinney cited the work of Nunnally and Bernstein who concluded for exploratory work, Cronbach's (1951) alpha values of .7 or higher are considered acceptable (as cited in McKinney, 2007, p. 90). Additionally, McKinney utilized multiple regression models to examine the degree to which each independent variable predicted or explained the dependent variable (p. 90). Both simultaneous multiple regressions and hierarchical multiple regression models were used (pp. 91-93).

In a discussion of reliability and validity, McKinney (2007) cited the acceptable nature of the reform documents, through review and refinement of KDE, as acceptable for the study (p. 97). Also, McKinney noted the attentive nature by which audit team members were trained for the Scholastic Audit process and, thus, enhancing external reliability (pp. 97-98). For a more specific analysis of validity, McKinney deferred to the state documents as a logical extension for validity since they were the "official policy of

the Commonwealth” (pp. 98-99).

In the results section, McKinney (2007) detailed the findings of each of his four empirical research questions (p. 111). For Research Question 1, McKinney posed the following:

To what degree does Leadership (Standard 7) affect Curriculum (Standard 1), Instruction (Standard 3), and the Academic Index? (p. 114)

McKinney found that Leadership had a significant impact on Curriculum, $F(1, 179) = 98.33, p < .001$, with an Adjusted R^2 of .35. As for the influence of Leadership on Instruction, McKinney also noted this relationship to be significant, $F(1, 179) = 100.81, p < .001$, with an Adjusted R^2 of .36 (p. 114). Through assessing the raw coefficient b , both Curriculum and Instruction would increase by .6, with a one unit increase in Leadership. McKinney discovered that Leadership was also significant on the Academic Index, $F(1, 179) = 62.74, p < .001$ (p. 115). Lastly, McKinney noted that the effect size of Leadership at .26 was less than the impact of Leadership on Curriculum and Instruction.

McKinney (2007) reported his findings for Research Question 2:

To what degree do Curriculum (Standard 1) and Instruction (Standard 3) affect the Academic Index? (pp. 115-116)

Examining the impact of Curriculum and Instruction on the Academic Index, McKinney found the ANOVA significant, $F(2, 178) = 51.31, p < .001$, with an effect size of .36 (p. 116). While Instruction is statistically significant at the $p < .001$ level, Curriculum was close to but not significant, with $p = .053$. In McKinney’s model, Instruction is found to be more impactful than Curriculum (p. 116).

For Research Question 3, McKinney (2007) asked the following:

To what degree do Demographic factors affect Leadership (Standard 7),

Curriculum (Standard 1), Instruction (Standard 3), and the Academic Index? (p. 116)

Of the eight demographic variables utilized by McKinney in the multiple regression model, only Percent Free and Reduced Lunch yielded a significant relationship for Standard 7, Leadership (p. 117). Overall, the ANOVA model was significant between the demographic variables and Leadership, $F(8, 172) = 3.72, p < .001$, with an Adjusted R^2 of .11—a small effect (p. 117). For the regression model of Standard 1, Curriculum on the demographic control variables, only Year of Audit and Percent Free and Reduced Lunch produced significance, $F(8, 172) = 2.956, p = .004$, with an Adjusted R^2 of .08 (p. 118). For the impact of demographic variables on Standard 3, Instruction, McKinney noted the significance of the model, $F(8, 172) = 5.141, p < .001$, with a small Adjusted R^2 of .16 (p. 119). Finally, the regression of the Academic Index on the demographics was also significant, $F(8, 172) = 34.16, p < .001$, with an extremely impactful Adjusted R^2 value of .60 (p. 120). The following demographic variables yielded significance on the Academic Index: Percent White, Percent Free and Reduced Lunch, County/Independent district, Percent Gifted, and Year of Audit—with Free and Reduced Lunch and Year of Audit having the strongest effects (p. 120).

For the final Research Question, McKinney (2007) posed the following:

To what degree do Curriculum (Standard 1) and Instruction (Standard 3) mediate the effect of Leadership (Standard 7) on the Academic Index, controlling for demographics? (p. 121)

For Research Question 4, McKinney conducted a three-step hierarchical multiple regression analysis. In Step 1, only demographic variables that were significant from the Research Question regressions were entered. Similarly, only variables significant from

regressions for Research Questions 1 and 2 were entered in Steps 2 and 3, respectively (pp. 121-122). During Step 1, the hierarchical multiple regression produced the following, $F(5, 175) = 54.88, p < .001$, with a strong Adjusted R^2 value of .60 on the Academic Index (p. 123). Percent White, Percent Free and Reduced Lunch, Percent Gifted, County/Independent district, and Year of Audit were all significant at the $p < .05$ level. During Step 2, Standard 7 (Leadership) was added and found to be significant to the model, $F(6, 174) = 62.38, p < .001$, producing an additional Adjusted R^2 of .07, from .60 to .67 (p. 123). Also, McKinney conducted a full and reduced comparison F test, $F(1, 174) = 39.50, p < .001$, confirming the significance of this step (pp. 123-124). Lastly, during Step 3, McKinney added Standard 1 (Curriculum) and Standard 3 (Instruction) to observe how they mediate for Leadership, after accounting for demographic control variables: $F(8, 172) = 56.50, p < .001$, an additional R^2 of .04, moving the Adjusted R^2 from .67 to .71 (p. 124). Again, McKinney confirmed Step 3 results with a full and reduced comparison F test, $F(2, 174) = 13.01, p < .001$. This last Step 3 of the hierarchical regression model explained 71% of the total variance on the Academic Index (p. 124). In essence, the last regression model confirmed the mediation of Leadership through the significant Standard Instruction, with Curriculum also close to significance at $p = .053$ (p. 124).

In summary, McKinney (2007) ascertained that through factor analysis that one factor could load for all of the indicators on each of the three standards: Standard 1 (Curriculum) with 56.9% of the variance explained; Standard 3 (Instruction) with 51% of the variance explained; and Standard 7 (Leadership) with 54.6% of the variance explained (p. 147). Additionally, Cronbach's alpha revealed a high degree of internal reliability for all three of the aforementioned standards (p. 147). McKinney noted, in response to Research Question 1, that leadership played an invaluable role in selecting and developing

the curriculum—which in turn had “profound consequences on achievement goals” (p. 149). In terms of the impact of leadership on instruction, McKinney found leadership influenced 36% of the overall instruction in a school and, thus, had an indirect impact on accountability scores. Additionally, an effective instructional leader could indeed improve accountability measures, with Leadership accounting directly for 26% of the variability on the Academic Index (p. 150).

In analyzing Research Question 2, McKinney (2007) stressed the paramount importance of instruction on accountability outcomes, with Instruction accounting for 36% of the overall variance on the Academic Index (p. 151). Particularly important, McKinney noted the importance of these findings, since they incorporated a macro-level view of statewide data (p. 151). For analysis of Research Question 3, McKinney noted the profound impact of demographics on accountability measures. Interestingly, the researcher proffered that the slightly negative impact of demographic control variables on Leadership could have stemmed from high-poverty, low-performing schools being paired with ineffective leaders—or more effective leaders leaving these high-poverty, low-performing schools (pp. 152-153). While demographic factors accounted for little variance in Curriculum and Instruction, the impact of demographic variables on the Academic Index was extremely noteworthy, explaining nearly 60% of the overall variance (pp. 153-154). Analyzing the final research question, McKinney again stressed the impact of demographics on achievement—with noticeable gaps between whites and other races continuing to appear (pp. 155-156). Overall, the researcher noted that while Leadership and Instruction only account for 7% and 4%, respectively, of the overall variance on the Academic Index, after the demographic control variables are removed, these two factors were alterable, consistently in the public and political eye, and could be effectively

developed by schools and districts (p. 158).

In essence, McKinney (2007) delivered a solid foundational research guide for which to build the current set of dissertations regarding the *Standards and Indicators for School Improvement* and the Scholastic Audit process—a seminal reform document and reform measure that provided a wealth of analysis opportunities for Kentucky educators. Aptly stated by McKinney, “a significant role of accountability (just beginning to unfold in earnest in Kentucky) is that it tends to uncover deep issues that could remain hidden in the more traditional bureaucratic culture” (p. 174). Through the current set of dissertations, the researchers are not only analyzing this accountability culture, but, perhaps even to a greater degree, providing further evidence for vested educational parties (principals, teachers, students) to examine for their actual empirical progress.

Extending the prior work of McKinney (2007), B. C. Ennis (2007) examined the effectiveness of professional development modules, pertaining to instructional leadership through Kentucky’s accountability model. With the *Standards and Indicators for School Improvement (SISI)* as the guiding document of the Scholastic Audits, B. C. Ennis specifically focused on Standard 6 (Professional Growth, Development and Evaluation) and Standard 7 (Leadership) and the subsequent impact on Kentucky’s achievement outcomes.

For the study, B. C. Ennis (2007) employed the following central research question: “How is Leadership (Standard 7) related to accountability outcomes (The Academic Index), controlling for the demographic context of the schools and as mediated by Professional Growth, Development and Evaluation (Standard 6)?” (p. 12). B. C. Ennis conducted a quantitative study design with the school as the unit of analysis. Similar to the other dissertations in the series, B. C. Ennis accessed a secondary database from KDE that

contained Scholastic Audit data collected by Scholastic Audit teams from 181 audited schools ($N = 181$) among the five audit cycles of 2001-2005.

B. C. Ennis (2007) noted the following reasons for the significance of the study: (a) provided further research in connecting professional development and leadership to student achievement, in light of demographic factors; (b) gauged what attributes a principal must possess to enhance student outcomes and how professional development would augment these attributes; (c) developed further the efficacy of the SISI and school improvement efforts in Kentucky; (d) and offered more research on ways that educators could improve achievement (pp. 14-16). Limitations to generalizability included the state specific process of accountability and the SISI for Kentucky, only elementary teachers from the KDE secondary database were included in the study, data were collected from different Scholastic Audit teams over multiple years, only two specific SISI standards were examined, and the improvement efforts of Kentucky schools could have possibly been influenced by federal accountability measures that were not accounted for in the Scholastic Audit process (pp. 16-18).

For the methodology of the research, B. C. Ennis (2007) employed multiple regressions as the main form of analysis (p. 149). The dependent variable was the Academic Index and the alterable independent variables were Standard 7 (Leadership) and Standard 6 (Professional Growth, Development and Evaluation) from the SISI document. Additionally, B. C. Ennis used demographic factors as the control variables. The researcher confirmed the validity and reliability of SISI Standards 6 and 7 through factor analysis and Cronbach's alpha (p. 168). The multiple regressions examining relationships between these standards and the Academic Index aligned with external criterion evidence validity as proposed by Nitko (as cited in B. C. Ennis, 2007, p. 168).

For the factor analysis of Leadership, a single factor loaded for this Standard 7. This single factor accounted for 54.6% of the total variance among the indicators. For Standard 6 and its 12 indicators, the factor analysis revealed two factors with eight and four indicators loading on the two substandards Professional Development and Professional Growth & Evaluation. KDE had previously specified the two factors with six indicators each. B. C. Ennis (2007) determined a Cronbach's coefficient alpha value of .915 for Standard 7, confirming the factor analysis. For Standard 6 analysis, B. C. Ennis also discovered a high degree of internal reliability, confirming the factor analysis of Standard 6 with Cronbach's coefficients of .904 and .825 for Components 1 and 2, respectively. B. C. Ennis employed a hierarchical regression model that provided a .70 effect size and, thus, indicating a strong effect on Leadership as mediated by Component 1, Professional Development but not for Component 2, Professional Growth and Evaluation (p. 195).

For Research Question 1, "To what degree do Demographic Factors affect Leadership (Standard 7), Professional Growth, Development and Evaluation (Standard 6), and the Academic Index?" B. C. Ennis (2007) stated the study added significantly to the field of research between the relationship of demographic factors and student achievement (p. 205). B. C. Ennis confirmed that only Free and Reduced Lunch students were significant for both Standard 6 and Standard 7 from the SISI, with beta values decreasing as the number of Percent Free and Reduced Lunch students increased. Additionally, the Year of Audit was significant for Component 1 of Standard 6, Professional Development (p. 205). B. C. Ennis proffered that the negative effect on Leadership and Professional Development due to the increased number of Free and Reduced Lunch students attributed to the amount of time spent working with this specific population, which subsequently

detracted from school improvement efforts (pp. 205-206). Also, B. C. Ennis noted, in regards to Variance Points 6.2b and 7.1c from the Eleven Common Variance Points, that low-performing schools did not afford educators the same resources to ensure school success through reform efforts as the high-performing schools and, thus, significantly and negatively impacted Leadership and Professional Development (p. 206).

For Research Question 2, “To what degree does Leadership (Standard 7) affect Professional Growth, Development and Evaluation (Standard 6) and the Academic Index?” B. C. Ennis (2007) discovered that leadership significantly impacts all of the aforementioned, including Component 1 (Professional Development) and Component 2 (Professional Growth and Evaluation) of Standard 6, at the $p < .001$ level. Leadership accounted for 49% of the variance in Professional Development, 47% of the variance in Professional Growth and Evaluation, and 25% of the variance on the Academic Index (p. 208). B. C. Ennis noted “schools will not improve unless the people in the schools also improve” (p. 208). With schools being judged primarily on student outcomes, B. C. Ennis argued that it was imperative that schools develop outstanding teachers and administrators (p. 209). The B. C. Ennis study provided further empirical confirmation, in addition to McKinney (2007), that leadership does indeed influence student achievement (B. C. Ennis, 2007, p. 209).

In terms of Research Question 3, “To what degree does Professional Growth, Development and Evaluation (Standard 6) affect the Academic Index?” B. C. Ennis (2007) used simultaneous regression to analyze the impact. Although KDE divided the 12 indicators of Standard 6 into two equal groups of six indicators, B. C. Ennis found that in fact, after conducting factor analysis, that Professional Development (Component 1) loaded with eight factors and Professional Growth and Evaluation (Component 2) loaded

with only four factors (pp. 209-210). B. C. Ennis discovered that Professional Development had a profound impact on student achievement, thus adding to the literature regarding the connection between professional development and enhanced student outcomes. Contrastingly, Professional Growth and Evaluation had no significance on student achievement.

For Research Question 4, “To what degree does Leadership (Standard 7) affect accountability outcomes (the Academic Index), controlling for the demographic context of the schools and as mediated by Professional Growth, Development and Evaluation (Standard 6)?” B. C. Ennis (2007) used hierarchical multiple regression. When determining whether adding more variables improved the model from Step 1 (Demographics) to Step 2 (Demographics and Leadership) to Step 3 (Demographics; Leaders; and Professional Growth, Development and Evaluation), B. C. Ennis employed a full and reduced model *F* test (p. 211). For all of the aforementioned steps, Percent White, Percent Gifted, Percent Free and Reduced Lunch, and Year of Audit all were statistically significant at the $p < .001$ level (p. 211). Moreover, two demographic factors, Percent Limited English Proficient and County or Independent schools were statistically significant at the $p < .05$ level (p. 211).

Demographic factors accounted for 61% of the variance in Step 1, with 7% added by Leadership and 2% added by Professional Growth, Development and Evaluation in Step 2 and Step 3, respectively (pp. 211-212)—confirming the earlier findings of McKinney (2007). The two demographic factors with the most significance—Percent Free and Reduced Lunch and Year of Audit—shed light first on the difficulty of improving high-poverty, low performing schools and second on the fact that the scores of the audited schools were improving from 2001 to 2005, evidence that the KERA comprehensive

reform model was working effectively throughout the Commonwealth—scaling up overall the dimension of excellence (B. C. Ennis, 2007, pp. 212-213). However, B. C. Ennis lamented that the same overall gains with Year of Audit had not been reflected in the other primary dimension, equity, where other demographic factors and achievement gaps had not been reduced—with the significance of Percent White and poverty confirming the notion of persistent achievement gaps (p. 213).

In summary, B. C. Ennis (2007) determined that Professional Development did indeed play a strong role in enhancing academic outcomes through Leadership. However, as B. C. Ennis quipped near the end of her discussion of results, knowing what to do and actually doing it are two different constructs: “Diagnosis is important. However, the cure has not been prescribed for academic achievement for all schools. Nor are the resources to effect the cure always adequate” (p. 215).

Regarding the next dissertation in the current set of studies implementing the *Standards and Indicators for School Improvement (SISI)* and the Scholastic Audit process, Saravia (2008) used secondary data from the KDE Scholastic Audit database to analyze the impact of the Learning Environment (SISI Standard 4 and 5) and the Academic Index, while controlling for demographic factors (p. vii). More specifically, Saravia assessed the impact of school culture and parental involvement as two important standards in the set of three that comprise the larger Learning Environment of the SISI to determine their impact on elementary school student achievement in Kentucky (p. 22).

Saravia (2008) used the following central research question for the study: “Within the context of demographic factors, what are the interrelationships between school culture and parent involvement with respect to academic achievement?” (p. 14). For the study, Saravia noted the following areas of significance: (a) the level of school culture and parent

involvement impact on student achievement; (b) the degree to which schools can understand and make accommodations for those variables that can be controlled and those variables that cannot; (c) the ways school culture and parent involvement mutually impact student achievement, and how they change in the presence of certain demographic factors; (d) and further research to extend the previous works of B. C. Ennis (2007) and McKinney (2007).

In terms of limitations, Saravia (2008) noted the fluctuating composition of Scholastic Audit teams as a possible limitation. Second, Saravia found that the self-reporting of audit data and the perceptions of Audit teams in collecting evidence could have posed as limitations. Third, the secondary database of the Scholastic Audit data and the secondary database of demographics (the Kentucky Performance Report) were both compiled by KDE and, thus, outside the control of the researcher. Fourth, the voluntary and non-voluntary nature of the Scholastic Audit process could have also influenced the willingness of schools to participate and, thus, affected the accuracy of the data. Fifth, potential misreporting of the Free and Reduced Lunch percentage by parents of students could have skewed demographic data. Sixth, the Scholastic Audit reports assumed interval spacing and did not accurately reflect the “more extreme gap” between Levels 2 and 3, as opposed to Levels 1 and 2 and Levels 3 and 4. Seventh, there was not an equal representation in the sample of the three levels of schools: assistance schools, schools in progress, and schools that met the goal. And, finally, generalizability would have been limited since this study only includes data from a Scholastic Audit process that was respective to Kentucky (pp. 18-20).

Saravia (2008) used quantitative analysis to explore relationships among SISI Learning Environment standards (Standard 4 and Standard 5), controlling for demographic

factors, while using data from a secondary database of Scholastic Audit data, maintained by KDE (p. 109). Saravia focused specifically on elementary schools with a designation of P-4, P-5, and P-6—since the accountability measures for elementary schools occurred during the third, fourth, and fifth grades—with third grade accounting for the Norm-Referenced Test, and fourth and fifth grades accounting for the Academic Index (p. 117). The sample for the study was 181 audited elementary schools ($N = 181$) during five rounds of Scholastic Audits from 2001-2005 (p. 117). The unit of analysis for the study was the school. In terms of the dependent variable, Saravia used the Academic Index (AI), which was the major component of the overall CATS Accountability Index (p. 117). For independent variables, Saravia used the following two: (a) control variables of demographic factors and (b) alterable SISI Standard 4 (School Culture) and Standard 5 (Student, Family and Community Support). Trained Scholastic Audit teams collected evidence during the audits that constituted the data entered for Standards 4 and 5 (p. 118).

For descriptive statistics, Saravia (2008) reported both the dependent and independent variables (p. 124). For psychometric analysis, Saravia used factor analysis to determine whether a single score could be reported for each standard, thus reducing the 11 descriptors for Standard 4 and five descriptors for Standard 5 of the SISI into single constructs (pp. 124-125). The researcher determined that Cronbach's alpha for each of the variables was well above the 0.7 value considered to be minimal scale reliability (p. 125). Saravia also used multiple regression analyses, both simultaneous and hierarchical, to analyze the secondary data (pp. 127-128). Finally, Saravia employed inferential tests focusing on β , R^2 , and ΔR^2 during multiple regressions (p. 129).

In terms of validity, Saravia (2008) noted that prior to the dissertations conducted by the students of Drs. Miller and Smith, KDE had done little, outside of Koger and

Thacker (as cited in Saravia, 2008, p. 133), to confirm the validity and reliability of the Scholastic Audit process, from whence the data originated. Also, being dependent on secondary data, Saravia noted the impossibility of assessing the validity and reliability of the original data and that the researcher must trust the Scholastic Audit processes (p. 134).

Through a discussion of results, Saravia (2008) explored the findings of his three empirical research questions. Saravia discovered a strong correlation between SISI Standard 4 and Standard 5 ($r = .795$) and several less impactful correlations among other variables (between .4 and .7) (p. 145). Saravia reported the greatest influence on the AI was Standard 5 (Student, Family and Community Support) with a $r = .588$. Confirming the previous work of B. C. Ennis (2007) and McKinney (2007), Saravia (2008) found that Free and Reduced Lunch percentages and the Year of Audit had sizeable negative ($r = -.581$) and positive ($r = .475$) relationships, respectively, with the Academic Index (p. 145).

Saravia (2008) posed the following for Research Question 1:

To what degree do:

Demographic factors (DEMV) influence the Learning Environment

- a. Standard 4, School Culture (SCULT)?
- b. Standard 5, Student, Family and Community Support (SFCS)?

and the Academic Index (AI)? (p. 148)

In answering Research Question 1, Saravia found demographic factors greatly influenced Standard 4 (School Culture), $F(13, 167) = 3.791, p < .001$, with an Adjusted R^2 of .168. Saravia explained that nearly 17% of the variation in School Culture could be explained by demographic factors (p. 149). Extended School Services was the only demographic variable that was significant with $p = .05$ (p. 149). Overall, a simultaneous multiple regression found statistical significance with demographic factors on Standard 5 (Student,

Family and Community Support) with $F(13, 167) = 4.606, p < .001$, with an effect size of .207 (p. 150). Additionally, for Standard 5 only one variable, free and reduced price lunch participation, was statistically significant with $p = .03$ (p. 150). Finally, for Research Question 1, demographic factors were also significant on the AI, with $F(13, 167) = 23.40, p < .001$, explaining nearly 62% of the variance in the AI (p. 151).

Saravia (2008) listed the following for Research Question 2:

To what degree does:

The Learning Environment

- a. Standard 4, School Culture (SCULT)
- b. Standard 5, Student, Family and Community Support (SFCS)

influence the Academic Index (AI)? (p. 153)

The result of the simultaneous multiple regression was significant with $F(2, 178) = 52.964, p < .001$, with an Adjusted R^2 of .366 (p. 153). In terms of unstandardized raw b coefficients, one unit on the School Culture scale increased the AI by 3.339 units, and the Student, Family and Community Support yielded an increase of 4.517 units.

Finally, Saravia (2008) found statistically significant results for the following Research Question 3:

To what degree does:

The Learning Environment

- a. Standard 4, School Culture (SCULT)
- b. Standard 5, Student, Family and Community Support (SFCS)

influence the Academic Index (AI), controlling for the demographic factors (DEMV)? (p. 154)

Saravia employed a two-step hierarchical multiple regression to control for the

demographic factors that were still significant. In Step 1, the ANOVA was statistically significant with $F(7, 173) = 42.599, p < .001$, with an Adjusted R^2 of .618 (p. 154). Saravia noted all remaining demographic variables as statistically significant except Extended School Services (p. 154). In Step 2, the ANOVA was significant with $F(9, 171) = 54.234, p < .001$, with an Adjusted R^2 of .727 (p. 155). Again, all remaining demographic variables were statistically significant, save Extended School Services (p. 155). Overall, after accounting for demographic variables, School Culture (Standard 4) and Student, Family and Community Support (Standard 5) contributed 11% additional variance to the AI, with an effect size of .727 (p. 155).

In summary, the research of Saravia (2008) again confirmed (see B. C. Ennis, 2007; McKinney, 2007) the feasibility of the *Standards and Indicators for School Improvement (SISI)* as a whole school reform tool (Saravia, 2008, p. 182). More specifically, both School Culture (Standard 4) and Student, Family and Community Support (Standard 5) yielded statistically significant impacts on student achievement through the Academic Index, with effect sizes of .274 and .370, respectively. Much like McKinney (2007), Saravia (2008) concluded his research by questioning the commitment of Kentucky to contribute the necessary funding to change high-poverty, low-performing schools—not only with additional resources but in the way schools are perceived by its citizenry (p. 194).

Building upon the set of previous dissertations of B. C. Ennis (2007), McKinney (2007), and Saravia (2008) on elementary schools, Todd (2010) examined the secondary database of KDE Scholastic Audit data from 83 Kentucky high schools ($N = 83$), during the same five rounds of Scholastic Audits between 2001 and 2005, to research the direct and indirect effects of leadership on student achievement (p. v; p. 120). For the purpose of

the research, Todd noted the research would add to the literature for the impact of leadership on school outcomes in curriculum and instruction at the secondary level and to examine relationships among demographic factors and student achievement (p. 13).

Todd (2010) cited the following central research question: “Based on Scholastic Audits, what are the effects of instructional leadership on Kentucky accountability scores, as mediated by curriculum and instruction?” (p. 14). In terms of the significance of the study, Todd detailed the following six distinct ways that the research would add to the field, specifically the opportunity: (a) to examine instructional leadership in the context of the Kentucky Education Reform Act of 1990; (b) to examine instructional leadership in the first state to move standards-based reform to whole school reform through the *Standards and Indicators for School Improvement* (SISI); (c) to examine instructional leadership through a secondary database of Scholastic Audit data; (d) to supplement the gap in literature on mediated leadership at the high school level; (e) to supplement the current set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008) and uniquely add to these studies through the lens of the high school level; and (f) to extend the work of McKinney (2007), but at the high school rather than the elementary level (Todd, 2010, pp. 17-18).

For limitations, Todd (2010) noted several ways the research could be influenced and limited in generalizability. First, the SISI was a Kentucky specific document and, thus, might not extend to other states in terms of education reform (p. 19). Second, academic accountability is also state specific and, thus, could limit generalizability. Third, the results of the state tests through the Kentucky Reform Act of 1990 only apply to Kentucky, with each state penning different assessment measures. Fourth, the research of Todd was limited to high schools only. Fifth, the reliability of the Scholastic Audit teams,

which collected the evidence-based data for the Scholastic Audits, had changing compositions to team structure from year to year and school to school, thus limiting reliability. Sixth, the year of audit presented certain limitations through the reliability of fluctuating Scholastic Audit teams and the natural turn-over of teachers and administrators in schools. Finally, the research of Todd coincided with certain federal provisions (No Child Left Behind Act of 2001) that influenced Kentucky education reform during this time period (p. 19).

Todd (2010) employed quantitative research through an analysis of a secondary database maintained by the Kentucky Department of Education (KDE) to assess the impact of instructional leadership on curriculum and instruction through Academic Index scores in Kentucky high schools and to assess the influence of demographic factors on student achievement (p. 85). Todd used multiple regression as the main form of analysis, with the school as the unit of analysis. For the dependent variable, Todd selected the Academic Index as the lens for examining student achievement in Kentucky high schools. For the independent variables, Todd used three standards from the SISI: Standard 7 (Leadership) as the alterable independent variable, and Standard 1 (Curriculum) and Standard 3 (Instruction) as the mediating independent variables (p. 86; pp. 97-103). Additionally, Todd assessed the influence of certain demographic control variables, pulled from the Kentucky Performance Report (KPR), to further assess the influence on the Academic Index (p. 86; pp. 97-99).

In terms of methodology, Todd (2010) used descriptive analysis, psychometric analysis, and multiple regression in the research (pp. 104-109). Psychometric analysis included the use of factor analysis and Cronbach's alpha (as cited in Todd, 2010, p. 105). For examining the four research questions, Todd cited multiple regression as a way to

assess data involving two or more independent variables in order to explain the influence of a single dependent variable (p. 106). Regarding multiple regression methodologies, Todd used two different regression techniques: simultaneous and hierarchical multiple regression (pp. 108-109).

For the reliability and validity of the study, Todd (2010) discussed the “scrutiny of validity and reliability for the CATS assessments,” but noted that this same skepticism by critics had not transferred to the Scholastic Audit process. Much like B. C. Ennis (2007) and Saravia (2008), Todd (2010) cited Koger and Thacker who conferred some reliability to the Scholastic Audit process through their observation of well-trained Scholastic Audit teams (as cited in Todd, 2010, pp. 112-113). Todd detailed the strong external criterion validity and psychometric quality of the SISI standards through the current set of aforementioned completed dissertations, led by Drs. Miller and Smith at Western Kentucky University, confirming the standards as a unified construct (p. 113). The work of Todd (2010) represented the first analysis of the SISI and Scholastic Audit process at the high school level and paralleled the work of McKinney (2007), who studied the elementary level.

Subsequently, Todd (2010) detailed the findings from each of the empirical research questions. Again, the data from the research were collected from 83 Kentucky high schools ($N = 83$), out of the 224 total high schools in Kentucky, between 2001 and 2005 (p. 119). Todd only analyzed one SISI Standard, Standard 7 (Leadership), from the Efficiency grouping (Standards 7, 8, and 9). For Research Questions 1-3, Todd (2010) utilized simultaneous multiple regression and used hierarchical multiple regression for Research Question 4 (p. 146).

Todd (2010) posed the following for Research Question 1:

To what degree do demographic factors affect Standard 7 (Leadership), Standard 1 (Curriculum), Standard 3 (Instruction), and the Academic Index in Kentucky high schools? (p. 147)

In examining the relationship of the independent variables of demographic factors and Standard 7 (Leadership), Todd noted significance with $F(9, 83) = 2.26, p < .001$, with none of the individual variables showing significance on Leadership (p. 148). The effect size for this regression had an Adjusted $R^2 = .12$. Thus, Todd gleaned that Leadership was independent of the influence of demographic control variables. When regressed on the demographic variables, Standard 1 (Curriculum) yielded $F(1, 83) = 51.15, p < .001$, with an Adjusted $R^2 = .008$ (p. 149). No independent demographic variables were statistically significant on Standard 1. Like Standard 7 (Leadership), Standard 1 (Curriculum) was essentially independent of the demographic variables (pp. 149-150). For Standard 3 (Instruction), the regression model found no significance of demographic variables: $F(9, 83) = 3.014, p < .001$, with an Adjusted $R^2 = .181$ (p. 150). Of note, Percent Free/Reduced Lunch had a beta coefficient of $-.395$ on Standard 3. Also, Year of Audit yielded a beta of $.281$ on Standard 3, with later years equating to “better quality instruction” (p. 151). Finally, demographics accounted for 65% of the variance on the Academic Index with significant ANOVA results, $F(9, 83) = 17.615, p < .001$, and an Adjusted R^2 value of $.65$ (p. 151). Among the demographic variables, Percent Free/Reduced Lunch and Year of Audit yielded the strongest influences with respective beta values of $-.38$ and $.46$ (p. 152).

For Research Question 2, Todd (2010) posed the following:

To what degree does Standard 7 (Leadership) affect Standard 1 (Curriculum), Standard 3 (Instruction), and the Academic Index in Kentucky high schools? (p. 153)

Todd discovered that Standard 7 (Leadership) had a significant influence on Standard 1 (Curriculum), with an ANOVA of $F(1, 83) = 51.147, p < .001$, and explaining an average variance of 40% on Standard 1 (p. 153). Leadership yielded a beta coefficient of .623 unit increase on Curriculum (pp. 153-154). Standard 7 also produced a significant impact on Standard 3 (Instruction): $F(1, 83) = 53.096, p < .001$, with an effect size of .39 (p. 154). Leadership consequently impacted Instruction with a beta coefficient of .63 (for Leadership). Finally, the Academic Index regression on Standard 7 (Leadership) was also significant, $F(1, 83) = 19.61, p < .001$, with an effect size of .20. Leadership impacted the Academic Index by a standardized beta of .44 (pp. 154-155).

Todd (2010) posed the following for Research Question 3:

To what degree do Standard 1 (Curriculum) and Standard 3 (Instruction) affect the Academic Index in Kentucky high schools? (p. 155)

The regression model for this research question examined the combined impact of Standard 1 and Standard 3 on student achievement, as measured through the Academic Index. The ANOVA model proved significant, $F((2, 83) = 26.01, p < .001$, with 39% of the variance on the Academic Index explained (p. 155). Of note, Todd (2010) stated that Standard 3 (Instruction) accounted for all of the influence in its combined impact with Standard 1 (Curriculum) on the Academic Index (p. 155).

Finally, Todd (2010) proffered the following as Research Question 4:

To what degree do Standard 1 (Curriculum) and Standard 3 (Instruction) mediate the effect of Standard 7 (Leadership) on the Academic Index, controlling for demographics in Kentucky high schools? (p. 156)

Todd utilized a hierarchical multiple regression model for Research Question 4. For the three separate steps, each of the ANOVAs for the hierarchical regression model were

significant. For Step 1, with demographics only, the findings of the ANOVA were significant, $F(9, 83) = 17.165, p < .001$, with an Adjusted R^2 value of .65 on the Academic Index (pp. 156-157). Additionally, the following demographic variables were significant at the $p < .05$ level during Step 1: County/Independent district, Percent Gifted, Percent Free/Reduced Lunch, Percent White, and Year of Audit (p. 157). For Step 2, Todd added Standard 7 (Leadership) to the model with significance, $F(10, 83) = 21.329, p < .001$, with a rise in the Adjusted R^2 value from .65 to .71 of the variance. In terms of demographics, Percent Gifted loses its status as statistically significant, as County/Independent district, Percent Free/Reduced Lunch, Percent White, and Year of Audit remained (p. 157). Todd noted that Leadership was also significant with a beta value of .284. Finally, for Step 3 of the hierarchical multiple regression model, Todd added Standard 1 (Curriculum) and Standard 3 (Instruction). Todd noted the significance of these additions to the model, $F(12, 83) = 22.019, p < .001$, with an additional variance of .05, from .71 to .76 in effect size (p. 157). Overall, after three steps of a hierarchical multiple regression model, 76% of the overall variance was explained on the Academic Index. The demographic factors yielding statistically significant beta values were Year of Audit, Percent White, Percent Free/Reduced Lunch, and County/Independent district (p. 157). Additionally, after accommodating for demographic control variables, both Standard 7 (Leadership) and Standard 3 (Instruction) were statistically significant with beta values of .215 and .279, respectively (p. 158). Interestingly, Todd noted the Instruction beta value exceeded that of Percent Free/Reduced Lunch, with Leadership nearly as high, and, thus, offering some hopeful research for educators that alterable factors can indeed make a significant impact on student achievement (p. 158).

In summary, Todd (2010) discovered that Instruction and Leadership played an

integral role through which Kentucky high schools, and specifically, administration can influence student achievement (p. 182). Furthermore, the work of Todd extended McKinney's (2007) elementary school research on the influence of leadership and instruction, yet at the high school level. Of note, the influence of Standard 1 (Curriculum) paled in comparison to Standard 3 (Instruction) and Standard 7 (Leadership), perhaps pointing to the fact that it matters less *what* you teach than how you teach it and who is leading the charge (p. 191). Furthermore, Todd suggested that the "guesswork" has been removed from curriculum in the academic reform equation with the inclusion of guiding documents such as "the Program of Studies, Academic Expectations, and Core Content" (p. 191). This also raises the possibility that there is less variation in curriculum (what to teach) because of the rigorous state guidelines and documents compared to instruction (how well is the content taught) because of differences in schools and teachers on this quality dimension. Aptly enough, Todd concluded her research with the quote "there is hope"—a hope that sound instruction and leadership can overcome the unalterable factors that students bring to school everyday (p. 201).

Summary

In the current review of literature, the researcher focused on the following pertinent areas: (a) foundational accountability, focused on the standards movement and achievement in Kentucky; (b) Academic Performance through the lens of the *Standards and Indicators for School Improvement* in Kentucky; (c) elementary science best practices in curriculum, classroom evaluation/assessment, and instruction; and (d) the confluence of Academic Performance, accountability, and science outcomes in Kentucky elementary schools.

Over the past three decades, since the seminal report *A Nation at Risk* (National

Commission on Excellence in Education, 1983), America has been in a period of exceptional educational reform. From accountability systems at the state, district, and school level, to standards-based reform, instruction, and assessment practices, to adopting punitive measures for schools and teachers not meeting the rigorous accountability demands, Kentucky has been firmly entrenched in both spearheading their own comprehensive system of whole school reform (Kentucky Education Reform Act of 1990) and following national regulations (No Child Left Behind Act of 2001). In terms of overarching accountability, the current review of literature incorporated the seminal work of Levin (1974) and his definition of accountability: “one [a system] that maximizes the utility or satisfaction of the relevant constituencies subject to resource (and possibly other) constraints” (p. 375). Or, perhaps even more essential (and succinct), Linn (2003) stated that accountability can be distilled into two questions that must be addressed: (a) what counts? and (b) who is held accountable? (p. 3).

Since the Kentucky Education Reform Act of 1990 (Kentucky Department of Education [KDE]), Kentucky educators observed a series of reform iterations pass through the state legislature in an attempt to answer generally the two questions above. After the first major reform movement during the aforementioned Kentucky Education Reform Act (KERA) of 1990, Kentucky successively enacted reform measures to meet the promises of KERA. The state legislature and authoring bodies of government grounded KERA in the following six learning goals, previously mentioned in the current review of literature: (a) able to use basic communication and mathematics skills for a variety of real-world situations; (b) able to develop and apply skills and abilities from the fields of mathematics, science, arts, humanities, practical living, and vocational training for a variety of real-world situations; (c) able to develop these skills to become self-sufficient learners; (d) able

to develop these abilities to become responsible members of the community, in both family and work; (e) able to develop these skills to think and solve real-world problems; and (f) able to develop these abilities to integrate and build upon past knowledge and learnings (KDE, 2012c). With pre-1990 Kentucky education marked with inequity and the all too common and sometimes not even subtle prejudice that all students could not learn at a high level, Steffy (1993) noted KERA was underscored with three guiding beliefs: that all students could learn at high levels, what is known in educational research could help Kentucky achieve this task, and what students learn should be consistent, regardless of whether a child lived in Pikeville or Paducah (pp. 6-7).

Several years after the enactment of KERA (KDE, 1990), it became apparent that the inclusion of words like reform and accountability and standards in state documents did not suddenly or magically cause them to appear in Kentucky classrooms. Thus, the Kentucky Department of Education (KDE) systematically began writing documents that in essence guided the reform movement for more than a decade. The *Standards and Indicators for School Improvement (SISI)* was the guiding document for whole school reform measures in Kentucky. Enacted from 703 Kentucky Administrative Regulation 5:120, the SISI delivered to Kentucky educators a framework for school improvement. Composed of three reform domains with three standards each, the SISI was divided into the following categories of standards: (a) Academic Performance with Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction); (b) Learning Environment with Standard 4 (School Culture), Standard 5 (Student, Family and Community Support), and Standard 6 (Professional Growth, Development and Evaluation); and Efficiency with Standard 7 (Leadership), Standard 8 (Organizational Structure and Resources), and Standard 9 (Comprehensive and Effective

Planning). Each standard was grounded in best practices indicators, with a total of 88 indicators across the 9 areas. (A complete list of standards and indicators from the SISI is located in Appendix A.) The current research only incorporates the Academic Performance standards, the first three standards, into the analysis. While the other standards are discussed through empirical articles or anecdotal reference, it is beyond the scope of this study to examine Standards 4 through 9.

In response to the need to reform low-performing schools across Kentucky, legislators passed KRS 158.6455 (Section 4) for teams of educators to conduct Scholastic Audits and Reviews at these low-performing schools, with the SISI as a means for identifying problem areas (Koger & Thacker, 2004, p. 1). Scholastic Audit and Review teams were composed of a defined combination of parents, teachers, school administrators, district administrators, university faculty representatives, and a Highly Skilled Educator. These trained Scholastic Audit teams subsequently audited a school, collecting evidence through the 4-point behavioral scale of implementation outlined in the SISI. While the vast majority of schools audited fell under the low-performing classification, several high-performing schools volunteered to be audited in order to ascertain what separated them from their not-as-successful counterparts—later helping to develop *variance points* (Lyons & Barnett, 2011, p. 1).

With an explication of accountability and Academic Performance through the current review of literature, the researcher then discussed the role of best practices in elementary science curriculum, classroom evaluation/assessment, and instruction, as aligned with SISI Academic Performance Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction). In examining the research behind sound practices in curriculum, classroom evaluation/assessment, and instruction, the

researcher gleaned certain trends that not only apply to elementary science, but, to a greater extent, could be generalized to all content areas across all grade levels.

First, since the early 2000s, the spread of Professional Learning Communities (PLC) transformed education over the past decade (DuFour et al., 2006). At the heart of professional learning communities, and its implications for curriculum, classroom evaluation/assessment, and instruction, is this idea that educators should focus on student learning. While not a revolutionary concept in theory, the application was much more dramatic. Through the lens of common assessments (“who gets it and who doesn’t”) of the standards, teachers were afforded the opportunity to cull the “mile wide, inch deep” curriculum of the past thirty years, hone in on students who had not mastered concepts, and use instructional practices focusing on what students do not know, rather than spending valuable time teaching what they already do know.

Second, the use of formative assessments also transformed not just science education, but assessment practices across every content. Assessments *for* learning, formative assessments are administered throughout a learning unit to gauge student knowledge and understanding before the final, summative assessment (DuFour et al., 2006, p. 55). Formative assessments alleviate the pressure of both teachers and students by affording a safety valve for understanding difficult concepts. If the goal is learning, then educators need a way to assess student knowledge (rather than predicting) before the end of a given unit.

Third, in order to meet the demands of scientifically literate students that can compete in a globalized society, another trend that emerged (again, aligned with this idea of a PLC) was vertical and horizontal communication regarding student learning. For many years, teachers operated in silos or, as the adage goes, “shut my door and I am

king/queen of my classroom.” However, this type of isolationism by educators was not in the best interests of the learners. Thus, through vertical and horizontal communication, teachers are afforded the opportunity to embed high-interest topics such as science across the curriculum through intensive learning units (Zemelman et al., 2012, p. 68).

At the nexus of Accountability, Academic Performance, and Science Outcomes, the current research seeks to push the field forward with an in-depth examination of how the Academic Performance standards—SISI Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction)—impacted the Academic Index on the Commonwealth Accountability Testing System (CATS), while controlling for demographic factors. The unit of analysis for this research is the school. From the numerous schools which were required to conduct a Scholastic Audit, this study focuses only on elementary schools ($N = 181$), and, specifically, science accountability scores, during the five testing cycles between 2001 and 2005. The conceptual framework for this study (see Figure 1) aligns with the theoretical framework, Systems Theory, selected by the researcher. With the litany of moveable parts that constitute the Kentucky reform movement of the 1990s and 2000s, systems theory afforded the researcher the opportunity to glean insight from the component parts and apply this knowledge to the larger whole of educational reform. Ultimately, at the intersection of Accountability, Academic Performance, and Science Outcomes, the goal of this study is to answer the following central research question: To what extent is Academic Performance (Standards 1-3 from Kentucky’s *Standards and Indicators for School Improvement*) related to elementary school science accountability scores?

CHAPTER III

METHODOLOGY

Introduction

As America grapples with preparing students for the uncertainty of the 21st century job market, one where science, technology, engineering, and math skills are at a premium to meet the highly technical, advanced professions of the future, and where critical thinking and collaboration are paramount, what happens in the classroom becomes even more significant as increased reform efforts focus on continuous improvement to achieve the burgeoning demands of education. In this study, the researcher examines how the first three standards in the *Standards and Indicators for School Improvement (SISI)*, collectively termed Academic Performance (Standard 1: Curriculum; Standard 2: Classroom Evaluation and Assessment; and, Standard 3: Instruction) affect the elementary science scores from the Academic Index of accountability testing in Kentucky.

Chapter III outlines the methodology of the study, detailing **Research Design, Definition of Terms, Description of the Data, The Accountability Process, Scholastic Audit Methodology, Description of the Variables, Analysis Plan, Empirical Research Questions, Reliability and Validity, and Ethical Issues**. The chapter concludes with a **Summary**.

Research Design

This study is a quantitative analysis of Scholastic Audit data from a secondary database maintained by the Kentucky Department of Education (KDE). The Scholastic

Audit was a means for evaluating the performance of a school. Through either a voluntary or mandated audit, the KDE assessed schools using the Scholastic Audit with the *Standards and Indicators for School Improvement (SISI)* instrument—an assessment tool composed of three sections with a total of nine standards (three standards per section). This study only examines the following: (a) any association between demographic factors (such as size, ethnicity, Appalachian status, and year of audit) and elementary science achievement scores; and (b) the Academic Performance Standards 1-3 (Curriculum, Classroom Evaluation and Assessment, and Instruction, respectively) of the SISI in relationship to the associated elementary science academic accountability outcomes for Kentucky—extrapolating the science scores from the Accountability Index from schools participating in a Scholastic Audit. As previously noted, for the current research, Academic Performance indicates the extent to which a classroom incorporates the components from the first three standards of the SISI and not as an indication of district and school achievement.

Achievement scores for the Academic Index, under the Accountability Index, can be linked to the student or the entire school. For this research, the school was the unit of analysis for collection of data by the Kentucky Department of Education (KDE) and for the study. The research represents a correlational design. Descriptive statistics, psychometric analysis, and multiple regression are performed with the data using SPSS 22.

Definition of Terms

The following list provides relevant definitions for each of the terms used to discuss Kentucky accountability in this research.

Academic Index for Elementary Schools (AI): The Academic Index constitutes 90.25% of the overall Accountability Index for schools with Norm-Referenced Testing

grades. Academic content values include: Reading, 19%; Mathematics, 19%; Science, 14.25%; Social Studies, 14.25%; On-Demand Writing Prompt, 2.85%; Writing Portfolio, 11.4%; Arts and Humanities, 4.75%; and, Practical Living/Vocational Studies, 4.75% (KDE, 2006a, Appendix D).

Academic Performance: Taken from the *Standards and Indicators for School Improvement* (KDE, 2003), Academic Performance referred directly to the first three Standards of the SISI document, collectively known as the Academic Performance standards (Standard 1, Curriculum; Standard 2, Classroom Evaluation/Assessment; and, Standard 3, Instruction), and encompassed those factors that could directly impact student learning at the classroom level. Therefore, the current research references Academic Performance in terms of its SISI meaning, as opposed to a final product of teacher and district performance measures.

Academic Performance Levels (NAPD Descriptions): The following four levels of performance descriptors were used to classify student academic achievement performance in each assessed content area of Commonwealth Accountability Testing System (CATS):

1. *Novice*: Student demonstrates minimal, limited, underdeveloped, and at times inaccurate content knowledge and reasoning; student communication is ineffective and lacks detail with no evidence of connections within or between content areas; and student uses strategies that are inappropriate.
2. *Apprentice*: Student demonstrates some basic content knowledge and reasoning ability; student communicates reasonably well but draws weak conclusions or only partially solves or describes; and student attempts appropriate strategies with limited success.
3. *Proficient*: Student demonstrates broad content knowledge and is able to

apply it; student communication is accurate, clear, and organized with relevant details and evidence; student uses appropriate strategies to solve problems and make decisions; and student demonstrates effective use of critical thinking skills.

4. *Distinguished*: Student demonstrates an in-depth, extensive, or comprehensive knowledge of content; student communication is complex, concise, and sophisticated with thorough support, explicit examples, evaluations, and justifications; student uses and consistently implements a variety of appropriate strategies; and student demonstrates insightful connections and reasoning. (KDE, 2006a, p. 57)

Assistance Line: “Represents the point below which a school becomes eligible for assistance from the state. The Assistance Line begins in 2002 at the baseline and ends in 2014 at 80” (KDE, 2006a, p. 2).

Baseline Accountability Index: The Accountability Index score was determined by averaging the performance during the 1998-1999 and 1999-2000 school years. Subsequent progress is measured against this index each biennium (KDE, 2006a).

Commonwealth Accountability Testing System (CATS): A testing system that replaced the Kentucky Instructional Results Information System (KIRIS) in 1998. Unlike its predecessor, CATS demonstrates a higher reliability with fewer performance assessments than its predecessor. CATS measured performance based on content area assessments, norm-referenced assessments, and additional data such as graduation rate, etc., compiled into an Accountability Index. Within this Accountability Index, the Academic Index was one subscore that reflected school performance on the criterion-referenced assessment portion, known as the Kentucky Core Content Tests (KCCT).

Goal Line: The goal line represents the straight linear path from the baseline accountability index (determined by 2000 CATS results) to the goal of 100 in 2014 (Lyons & Barnett, 2011).

Highly Skilled Educator (HSE): Highly Skilled Educators (HSE) are specially trained teachers and administrators who assisted schools and districts in the Assistance Level 3 sub-category (the lowest one-third of schools) of the “In Need of Assistance” category. HSEs worked with school faculty, staff, and students, focusing on the improvement of teaching and learning in the building as dictated by Kentucky’s SISI document (Legislative Research Commission, 2006).

Kentucky Core Content Tests (KCCT): The criterion-referenced tests administered in the spring of every school year to all students in Grades 4, 5, 7, 8, 10, 11, and 12. The KCCT represents the majority of a school’s accountability index under CATS. The KCCT is administered on the following grade level schedule: Reading in Grades 4, 7, and 10; Math in Grades 5, 8, and 11; Science in Grades 4, 7, and 11; Social Studies in Grades 5, 8, and 11; On-Demand Writing in Grades 4, 7, and 12; Arts and Humanities in Grades 5, 8, and 11; and, Practical Living and Vocational Studies in Grades 5, 8, and 10 (KDE, 2006a, pp. 4-5).

Kentucky Instructional Results Information System (KIRIS): A comprehensive measurement and assessment system, initiated by the Kentucky Education Reform Act (KERA) of 1990, that supported the statewide educational accountability system from 1991-1998; this system was replaced by CATS in 1998 (KDE, 2006a).

Leverage Point: An indicator that represented a distinguishing factor between Level 3 Assistance Schools, who failed to meet improvement goals as determined by the 2000-2001 Scholastic Audits, and schools who met these goals: with 17 of the 88

indicators designated as leverage points (KDE, 2003). These leverage points were “viewed as valuable in focusing the improvement efforts of schools and school districts across Kentucky” (Lyons & Barnett, 2011, p. 1).

Nonacademic Index: This index includes, at the elementary level, the nonacademic factors of attendance (3.8%) and retention (.95%), for schools with NRT grades. The Nonacademic Index is then calculated into the overall Accountability Index for elementary schools (KDE, 2006a).

Norm Referenced Test (NRT): The NRT component is based upon the Comprehensive Test of Basic Skills (CTBS/5) Survey. The mandated areas for the CTBS/5 include Reading, Language, and Mathematics. The NRT allows Kentucky students to be compared to other students across the nation. The KDE developed an “index” for comparing the weight of the 0 to 140 scale as the KCCT content areas to the National Percentile Ranges for the NRT:

Weight	National Percentile Range
0	1-24
60	25-49
100	50-74
140	75-99

The mean index score for students on the NRT is 5% of the Accountability Index for all school levels (KDE, 2004a).

Scholastic Audit: Guidelines were established under Section 4 of the Kentucky Revised Statutes (158.645) to review a “school’s learning environment, efficiency, and academic performance of students; evaluate each certified member assigned to the school; and, report to the Kentucky Board of Education about the school’s classification and more

specifically what would be required to improve teaching and learning in the audited school” (KDE, 2003; KDE, 2006a). Categories designed for the Scholastic Audit process include:

1. **Successful School:** As long as the Novice reduction and Dropout criteria have been satisfied, schools can be classified as a “Successful School” if the school is at or above its goal line. Also, a school could be considered a “Successful School” if the “school is in the Progressing area of the Growth Chart, and increased its Accountability Index in the second biennium. Also, the school can garner the “Successful School” identification if the school “passes any one of the five Recognition Points (i.e., 55, 66, 77, 88, 100).” (KDE, 2006a, pp. 2-16)
2. **Assistance School:** At the end of every biennium, CATS provides assistance for schools who do not meet performance goals, with scores below the assistance line. This set of schools is then divided into thirds, with the following three levels:
 - a. **Level 1 Assistance Schools** are the upper one-third of schools below the assistance line and “shall adhere to the requirements for a “Level 1” school as defined in 703 KAR 5:120 Section 2. Level 1 schools must conduct a scholastic review and self-study facilitated by the district’s professional development coordinator with assistance provided by Kentucky Department of Education staff. Assistance Level 1 schools may be eligible to receive Commonwealth school improvement funds.” (KDE, 2006a, pp.16-17)
 - b. **Level 2 Assistance Schools** are the middle one-third of schools below

the assistance line and “shall adhere to the requirements for a “Level 2” school as defined in 703 KAR 5:120 Section 3. Schools are required to receive a scholastic review by a team set up by KDE. The team must include local district members. Level 2 schools shall receive a scholastic review facilitated by a designee of the Commissioner of Education with assistance from the district’s central office staff. Assistance Level 2 schools may be eligible to receive Commonwealth school improvement funds.” (KDE, 2006a, pp. 16-17)

- c. Level 3 Assistance Schools are the lowest one-third of schools below the assistance line and “shall adhere to the requirements for a “Level 3” school as defined in 703 KAR 5:120 Sections 4, 5, 6, 7, 8, and 9. Schools will be scheduled for scholastic audits by an external team coordinated by KDE. Level 3 schools shall receive education assistance from a highly skilled educator under KRS 158.782 and a scholastic audit. Assistance Level 3 schools may be eligible to receive Commonwealth school improvement funds.” (KDE, 2006a, pp. 16-17)

Scholastic Review: The review provides an auditing process much like the Scholastic Audit. However, the audit team is scaled down in numbers with only two representatives from the regional service center and two to four representatives from the audited school’s district (KDE, 2003).

Standards and Indicators for School Improvement (SISI): The SISI document guides school improvement. The SISI lists 88 indicators under nine standards for school improvement, which are divided into the following three domains of best practices: (a) Academic Performance, (b) Learning Environment, and (c) Efficiency. The Academic

Performance domain includes Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction). The Learning Environment domain includes Standard 4 (School Culture), Standard 5 (Student, Family, and Community Support), and Standard 6 (Professional Growth, Development, and Evaluation). The Efficiency domain includes Standard 7 (Leadership), Standard 8 (Organizational Structure and Resources), and Standard 9 (Comprehensive and Effective Planning) (KDE, 2003). See Appendix A for a list of the nine standards in the SISI document.

Variance Point: Originally termed *Leverage Points*, referring to the SISI factors that distinguished high performing schools from schools failing to meet improvement goals, KDE changed the reference to *Variance Points* since these indicators “represented variance in audit results between the two groups of schools” (Lyons & Barnett, 2011, p. 1). These variance points helped Kentucky educators and school leaders in “prioritizing school improvement practices” (Lyons & Barnett, 2011, p. 1). KDE identified 27 indicators as variance points as a result of the 2002-2003 round of audits. When the 2000-2001 round of audits were compared to the 2002-2003 audits, six common variance points existed. Further, after the third round of audits during the 2004-2005 cycle, KDE found 11 Common Variance Points, under a revised criterion for statistical variance significance, for all three accountability cycles (2000-2001, 2002-2003, 2004-2005) (KDE, 2003).

Description of the Data

The audit teams, trained by the Kentucky Department of Education (KDE) with the *Standards and Indicators for School Improvement (SISI)* document as the primary assessment and evaluation instrument, collected the data through Scholastic Audits or Reviews for this study. One hundred thirty-one schools were either audited or reviewed during the 2000-2001 school year and, according to the KDE, were divided as follows: 83

elementary schools, 28 middle schools, and 20 high schools. Using the levels established by the KDE for Accountability Index scores, the following levels were determined: 47 Level 3 schools (requires Scholastic Audit), 50 Level 2 schools (requires Scholastic Review), 15 Level 1 schools (Voluntary Scholastic Reviews administered), and 18 successful schools (Scholastic Reviews only conducted in a sample of successful schools). During the 2002-2003 school year, 114 scholastic audits or reviews were conducted and given the following designations: 29 Level 3 schools, 32 Level 2 schools, 29 Level 1 schools, and 24 successful schools (KDE, 2003). During the 2004-2005 school year, 48 schools were classified as in need of assistance (KDE, 2004c). This study limited data to only those Kentucky elementary schools that completed a Scholastic Audit or Scholastic Review from 2000-2005. The data represent all the elementary schools that contained a P-4, P-5, and P-6 configuration. Since students before Grade 3 are not tested and included in the Academic Index, no P-3 elementary school configuration was included in this study. For this sample, $N = 181$ for the five rounds of audits conducted: with 2001 providing 56 cases, 2002 providing 17 cases, 2003 providing 80 cases, 2004 providing 5 cases, and 2005 providing 23 cases. Of note, 17 repeat incidents of audits of the same school during different testing cycles were removed to allow for the independence of the study design.

In addition to the Scholastic Audits, elementary science school scores from the Academic Index and specific demographic variables were required for this study. The school scores and demographic data were drawn from the Kentucky Performance Report (KPR) and merged with the Scholastic Audit data.

The Accountability Process

Under the Kentucky Education Reform Act (KERA) of 1990, schools were measured using accountability measures that assessed the overall performance of the

respective school. The Commonwealth Accountability Testing System (CATS) was that accountability measure for Kentucky in the late 1990s and early 2000s—replacing the Kentucky Instructional Results Information System (KIRIS) after the initial eight years of KERA. CATS used content area assessments, norm-referenced assessments, and additional results (e.g., retention and attendance at the elementary grade level) for a holistic measurement known as the Accountability Index.

One subscore of the Accountability Index was the Academic Index. The Academic Index reflected the performance of a school on the criterion-referenced tests termed the Kentucky Core Content Tests (KCCT) and the norm-referenced Comprehensive Test of Basic Skills (CTBS). The Kentucky Department of Education (KDE) mandated that all schools have an Accountability Index score of 100 by the year 2014. With its inception in 2000, the KDE divided these 14 years into seven two-year cycles (bienniums) to determine the pace toward meeting this goal. When calculating the Accountability Index, the Academic Index is used as one measure of this overall index. The Academic Index uses the following descriptors: (a) meeting goal, (b) progressing, or (c) needing assistance.

Under the CATS accountability system, each school determined by the KDE to need assistance went through the Scholastic Audit. Most schools who performed the Scholastic Audit were required by the KDE to do so—with only several high performing schools applying for the accountability measure. The audit not only used academic achievement as a measure of school health, but, also, learning climate, efficiency, leadership, and culture (Redding & Walberg, 2008). The researchers gleaned from the reform movement that it was not the students that lacked capacity (p. 72). Those schools labeled as “needing assistance” were then further sub-divided into three categories: the lowest one-third of those “needing assistance” were termed Level 3 schools and were made

to conduct a Scholastic Audit; the middle one-third of those “needing assistance” were labeled Level 2 and were given a Scholastic Review; and, the highest one-third of those “needing assistance” were called Level 1 and were asked to conduct an internal self-review (KDE, 2003).

The KDE developed the *Standards and Indicators for School Improvement (SISI)* to guide schools by developing protocols to pursue whole-school reform and improvement. As a way to measure the SISI, the KDE later developed the Scholastic Audit, a performance evaluation consisting of 88 indicators across nine standards. Each of the standards represented a different component of school reform. The nine standards are divided into three major categories: (a) Academic Performance, (b) Learning Environment, and (c) Efficiency. Each category contained three of the aforementioned standards. When conducting the scholastic audits, the evaluation team used a spreadsheet, divided into four columns, listing the nine SISI standards.

Scholastic Audit Methodology

The Division of School Improvement (Office of Leadership and School Improvement) of the Kentucky Department of Education (KDE) was principally responsible for spearheading the key procedures and administering the training for the Scholastic Audit and Scholastic Review teams. Scholastic Audit teams were composed of a parent, a teacher, a school administrator, a district administrator, a university faculty member, and a Highly Skilled Educator (HSE). Scholastic Review teams consisted of two representatives from the regional service center and two to four constituents from the “reviewed” school’s central office (KDE, 2003). The *Standards and Indicators for School Improvement (SISI)* were composed in 2000 as a guiding document to ground the Scholastic Audit process for public schools (Lyons & Barnett, 2011, p. 1).

While the SISI document served as the primary assessment and evaluation tool, the audit and review teams also gathered data from the following sources: (a) leadership and culture surveys; (b) school portfolio documents; (c) team experiences; (d) interviews; and, (e) observations (KDE, 2003). For the school portfolio, the school included the Comprehensive School Improvement Plan, state testing results, writing portfolio analysis data, non-academic data, district technology inventory, master schedule, school handbook information, school report card, School-based Decision Making Council records, teacher lesson plans, professional learning activities, and examples of student-generated work (p. 9). In addition to the aforementioned school portfolio artifacts, the school profile was developed through an amalgamation of data from classroom observations and formal and informal interviews with all vested school parties: students, teachers, principals, assistant principals, counselors, central office staff, and parents. The Scholastic Audit and Scholastic Review teams then analyzed all the collected data, comparing the data to the Performance Descriptors for each indicator of every standard under the SISI document.

For classification and categorization purposes, the review and audit teams used the following 4-point variable scale descriptions for each indicator:

Category 1—Little or no development and implementation

Category 2—Limited development or partial implementation

Category 3—Fully functioning and operational level of development and implementation

Category 4—Exemplary level of development and implementation. (KDE, 2004b, p. 3)

For this study, the 4-point variable scores for each indicator were converted to the scores of 0, 2, 5, and 7 (see McKinney, 2007; Saravia, 2008). The conversion of scores

reflected a more accurate picture for schools in the transition from the “limited development” of Category 2 to the “fully functioning and operational level of development” of Category 3. Also, the aforementioned scores conversion mirrored the arduous process of moving students from Apprentice to Proficient on the 140-point Kentucky Core Content Test scale (Novice, 0-39; Apprentice, 40-99; Proficient, 100-139; and, Distinguished, 140).

From the collected data, the audit and review teams made suggestions for improvement regarding each standard in a comprehensive report. Commendations were presented for outstanding performance in a given SISI standard indicator. The report was divided into the nine standards of the SISI, with the following three headings: Academic Performance (Standards 1, 2, and 3), Learning Environment (Standards 4, 5, and 6), and Efficiency (Standards 7, 8, and 9). For a full list of the nine standards, see Appendix A. While KDE recommends a comprehensive overview of all nine standards when reaching for whole school reform, the analysis of all nine standards are beyond the scope of this current study. Only the scores given for the indicators of the selected SISI Academic Performance Standards, Standard 1 (Curriculum), Standard 2 (Classroom Evaluation / Assessment), and Standard 3 (Instruction), are used in this research. Additional data purported by the Scholastic Audits and Scholastic Reviews may be used for additional research.

Description of the Variables

This section outlines the operational descriptions, coding, and grouping of variables. The Scholastic Audit served as the primary data source for this study, for Standards 1, 2, and 3. Other data were collected from the Kentucky Performance Report (KPR). Academic Index school scores and demographic variable information from the

KPR were combined with the Scholastic Audit data for the analysis of this research. The school is the unit of analysis ($N = 181$) for this current research. However, of the 181 schools audited during the observed time periods, 17 incidents were repeat audits and subsequently removed to maintain the independence of the study design.

Dependent Variable

Elementary science scores, extracted from the overall Academic Index (AI), functioned as the only dependent variable for this research. The AI is defined in the **Definition of Terms** section above. The AI from the CATS accountability scores composes 90.25% of the overall Accountability Index for elementary schools, with the Norm-Referenced Test (NRT) accounting for 5% and the Nonacademic Index (retention and attendance) accounting for 4.75% of the remaining overall Accountability Index (KDE, 2006a, Appendix D). Elementary science scores are measured through the administration of the Kentucky Core Content Test (KCCT) in fourth grade during the spring term. The KCCT for science contains six Open Response questions and 24 Multiple Choice items. Open Response items each contain a specific scoring rubric that gauges answers on a zero to four scale (KDE, 2006a, p. 5). Multiple choice responses are scored either a zero or one (incorrect or correct, respectively). In terms of comparing the weights of multiple choice and open response items, multiple choice accounts for one-third of the content area domain and open response for two-thirds. The Kentucky Department of Education uses these weights “to reflect the instructional importance of the open-response items and to provide item-mean scores (both school and state) that reflect the same weighting used in accountability calculations” (KDE, 2006a, p. 25). This current research only examines the elementary science scores from the Academic Index.

Independent Variables

This study includes two types of independent variables, in accordance with the conceptual model in Figure 1: (a) Control variables, the demographics, and (b) Alterable variables, Academic Performance. The Control factors align with Bloom's (1980) immutable, socio-demographic background. The Control factors reflect the previous work of McKinney (2007), B. C. Ennis (2007), Saravia (2008), and Todd (2010) on a larger sampling of demographic factors. The Alterable factors, according to Bloom (1980), can change as a result of the daily decisions of schools such as the implementation of instructional initiatives and curricular endeavors for teachers. The Alterable factors are Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction)—which combine to form the Academic Performance domain of the SISI document. Listed below are the operational definitions for each variable:

Control variables (DEMV). In examining the control variables, the researcher selected the following fourteen variables to test in this study. With the exception of Voluntary versus Involuntary status, the following fourteen variables have been used in prior studies of the *Standards and Indicators for School Improvement (SISI)* at the elementary level (see Saravia, 2008; B. C. Ennis, 2007). While only six of the control variables were found significant at the elementary level in previous studies, the researcher is including all fourteen variables in this study. Since the current study examines only the science scores from the Academic Index (constituting a major portion of the larger Accountability Index), there is the possibility a demographic influence may exist for the science accountability scores, different from the entire Academic Index, and is thus examined accordingly.

Size (SIZE): Number of students in Grades four. Grades four and five

students take the KCCT, which contributes to the elementary Accountability Index.

Grades three and six only take the NRT, which minimally contributes to the Accountability Index and provides no contribution to the dependent variable, the Academic Index. For the current study, only Grade four students were included in the demographic variable (SIZE) since fourth grade students are the only students who are assessed in the science component of the KCCT.

Ethnicity (%WHITE): Percentage of majority students (White) within each school, coded 1 = White, 0 = all other. These data are listed in the KPR after being reported by individual schools. No other ethnic groups are listed since the minority distribution in Kentucky is not statistically sufficient to create more categories.

Appalachian Status (APP): The Appalachian region of Kentucky has a distinct rural mindset, influenced by the history of mining for this area (Caudill, 1963). Nearly a third of Kentucky's students are labeled as Appalachian (Smith, 2005, 2006). This variable is coded 1 = Appalachian region, 0 = not Appalachian region.

% Free/Reduced Lunch (%FRED): The percentage of the student population receiving Free and Reduced Lunch prices (a continuous variable running from 1-100) provides an adequate approximation of the socioeconomic status (SES) of a school. While not the most accurate reflection of the SES, its accessibility makes the Free and Reduced Lunch percentage a feasible touchstone for many research studies (cf. Harwell & LeBeau, 2010). The Free and Reduced Lunch percentages is reported by the school to the Kentucky Department of Education (KDE) and then made available for viewing in the KPR.

Independent/County School Districts (CTY): Schools that are located in the county school districts/systems, as opposed to independent school districts. Fifty-five of the 175 Kentucky school districts are independent. This variable is dichotomously coded 1 =

County, 0 = Independent.

% Non-Disabled Students (%NDAB): The percentage of non-disabled students who receive no special education accommodations as reported by a school to KDE and then listed in the KPR—a continuous running variable from 1-100.

% Migrant Students (%MIG): The percentage of students participating in migrant programs, as reported by KPR as a continuous variable running from 1-100.

% Gifted/Talented Students (%GFT): The percentage of students participating in gifted and talented programs, as reported by KPR as a continuous variable running from 1-100.

% LEP Students (%LEP): A variable that reports the percentage of students enrolled in an English language learning class, where English is not their native language. This variable is a continuous variable running from 1-100, as reported by the KPR.

Year of Audit (YEAR): This variable reflects the year in which the Scholastic Audit was conducted. This variable was labeled as an interval variable from 2001 to 2005. Required to reach the Proficient level (100 or above) on the Accountability Index by 2014, schools were expected to make continuous growth each year (coded as an ordinal variable) based on the expected growth line, ascertained after determining the baseline of the school. The inclusion of the time marker, year of audit, is necessary to reflect the significant progress that the educational system of Kentucky has made under the Kentucky Education Reform Act (KERA) since its inception in 1990 (Alston et al., 1999; Petrosko, 2000; Pankratz & Petrosko, 2000; Haselton, 2004; Miller & Moore, 2006).

Voluntary vs. Involuntary (VOL): The voluntary versus involuntary variable represents the status for a school participating in a Scholastic Audit. Since a handful of high-performing schools volunteered to have an audit conducted, the inclusion of this

variable helps to account for any possible discrepancies in the measures of central tendency for the demographic impact of schools on KCCT elementary science scores.

% Title I (%TITLE1): The percentage of students who qualify for Title I funding, as reported by the KPR as a continuous running variable from 1-100.

% Female (%FEM): The percentage of students who are female. This variable is reported by the KPR as a continuous running variable from 1-100.

% Extended School Services (%ESS): The percentage of students who qualify for extended school services (ESS), as reported by the KPR as a continuous running variable from 1-100.

Academic Performance. The SISI document is divided into nine standards and grouped according to a specific domain. The full list of SISI standards is in Appendix A. In this study, the three alterable independent variables each come from the Academic Performance area: Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction). Standard 1, Curriculum, contains the following seven indicators (KDE, 2004c):

Curriculum (CURR)

(CURR1) There is evidence that the curriculum is aligned with the Academic Expectations, Core Content for Assessment, Transformations and the Program of Studies.

(CURR2) The district initiates and facilitates discussions among schools regarding curriculum standards to ensure they are clearly articulated across all levels (P – 12).

(CURR3) The district initiates and facilitates discussions between schools in the district in order to eliminate unnecessary overlaps and close gaps.

(CURR4) There is evidence of vertical communication with an intentional focus on

key curriculum transition points within grade configurations (e.g., from primary to middle and middle to high).

(CURR5) The school curriculum provides specific links to continuing education, life and career options.

(CURR6) There is in place a systematic process for monitoring, evaluating and reviewing the curriculum.

(CURR7) The curriculum provides access to a common academic core for all students. (KDE, 2004c, pp. 5-8)

The second alterable independent variable is Standard 2, Classroom Evaluation/Assessment. This Standard contains the following eight indicators:

Classroom Evaluation/Assessment (CEA)

(CEA1) Classroom assessments of student learning are frequent, rigorous and aligned with Kentucky's core content.

(CEA2) Teachers collaborate in the design of authentic assessment tasks aligned with core content subject matter.

(CEA3) Students can articulate the academic expectations in each class and know what is required to be proficient.

(CEA4) Test scores are used to identify curriculum gaps.

(CEA5) Multiple assessments are specifically designed to provide meaningful feedback on student learning for instructional purposes.

(CEA6) Performance standards are clearly communicated, evident in classrooms and observable in student work.

(CEA7) Implementation of the state-required Assessment and Accountability Program is coordinated by school and district leadership.

(CEA8) Samples of student work are analyzed to inform instruction, revise curriculum and pedagogy, and obtain information on student progress. (KDE, 2004c, pp. 9-12)

The third alterable independent variable is Standard 3, Instruction. This standard offers the following eight indicators:

Instruction (INST)

(INST1) There is evidence that effective and varied instructional strategies are used in all classrooms.

(INST2) Instructional strategies and learning activities are aligned with the district, school, and state learning goals and assessment expectations for student learning.

(INST3) Instructional strategies/activities are consistently monitored and aligned with the changing needs of a diverse student population to ensure various learning approaches and learning styles are addressed.

(INST4) Teachers demonstrate the content knowledge necessary to challenge and motivate students to high levels of learning.

(INST5) There is evidence that teachers incorporate the use of technology in their classrooms.

(INST6) Instructional resources (textbooks, supplemental reading, technology) are sufficient to effectively deliver the curriculum.

(INST7) Teachers examine and discuss student work collaboratively and use this information to inform their practice.

(INST8) There is evidence that homework is frequent and monitored and tied to instructional practice. (KDE, 2004c, pp. 13-16)

Analysis Plan

This section contains the following topics regarding the type of analysis used in this study: data checking/coding, descriptive statistics, psychometric analysis, and multiple regression.

Data Checking and Coding

Secondary data, collected by other researchers, constitute the type of information used in this research. The research of this study contains secondary data on Standard 1, Standard 2, and Standard 3, from the *Standards and Indicators for School Improvement* (SISI) document, retrieved from the Kentucky Department of Education (KDE) database after the administration of Scholastic Audits and Scholastic Reviews at designated school sites by trained teams of educators. Other data, such as the control variables of demographics, were collected from the *Kentucky Performance Report* (KPR).

According to Gorard (2002), the inclusion of secondary data in educational research offers several advantages. Gorard offers that while many academic institutions place a premium on “originality” for dissertation work, a researcher can use secondary data and still conduct original work (p. 234). Perhaps, according to Gorard, the most obvious advantages of using secondary data is speed and cost (p. 234). Sautter (2014) also purports “the major advantage of this research model [secondary analysis] is that it eliminates the time and expense of gathering data and relies on high-quality, reliable data collected by experts” (p. 24). Further, secondary data can lay the foundation for a researcher to argue that more research is needed based on the deficiencies with the current data (Gorard, 2002, p. 234).

In addition to advantages, the use of secondary data includes several disadvantages. For example, one problematic element of using secondary data arises when the available

data do not directly address the empirical research question (Cheng, 2014, p. 374). More specifically, in an attempt to frame the research question around the data, the researcher, frustrated with the existing data, tries to connect the research question to “fit” the current data (Sautter, 2014, p. 28). The secondary data provided by the KDE database in the present study avoids the aforementioned disadvantage as it contains the exact types of data requisite for addressing the empirical research questions. Another disadvantage of using secondary data includes missing or incomplete data sets. The current KDE secondary database, from which the present study extracted data, contains no missing or incomplete data since the SISI data compiled by the Scholastic Audit and Scholastic Review teams were complete in the collection process. Also, the demographic variables, listed in the KPR, contain no missing or incomplete elements since KDE mandated the submission of this data from each school in Kentucky as a part of the accountability procedures. Any possible inaccurate or incorrect data sets therefore were the result of the human errors in collection and transmission. Other issues of validity and reliability are discussed later in this chapter.

Descriptive Statistics

Descriptive statistics are calculated for the independent variables, which includes all the demographic control measures. Also, descriptive statistics are reported for the dependent variable of CATS results from the elementary school science scores on the Kentucky Core Content Test (KCCT). In regards to the Academic Performance variables for the *Standards and Indicators for School Improvement* (SISI), Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction), descriptive statistics are found in the **Psychometric Analysis** section.

Psychometric Analysis

In determining the extent to which data reduction is possible in the psychometric analysis of variables, factor analysis (FA) and Cronbach's (1951) alpha are calculated and conducted, respectively. Factor analysis is a form of psychometric calculation that allows the researcher "to reduce a large number of variables to a smaller number of variables, or factors..." through "finding patterns among the variations in the values of several variables..." (Vogt, 1999, pp. 107-108). Vogt defines a factor as "a cluster of highly intercorrelated variables" (p. 108). Variables that can subsequently be grouped together are therefore treated as one variable. Cronbach's (1951) alpha is a way to measure the internal reliability of a set of variables in an index (Vogt, 1999, p. 64). Measured from 0 to 1.0, Cronbach's (1951) alpha, as it approaches 1.0, purports a greater reliability that items in an index are measuring the same thing—with items possessing a Cronbach's (1951) alpha of 0.7 or above having a high level of internal reliability (Vogt, 1999, p. 64).

When applied to the current research, factor analysis and Cronbach's (1951) alpha helped determine whether the respective indicators for each of the three Academic Performance standards (Standards 1-3), from the *Standards and Indicators for School Improvement* (SISI) document, should be grouped together or analyzed as individual indicators. Considering the large number of indicators for each SISI standard, factor analysis and Cronbach's (1951) alpha were performed on each of the three SISI Academic Performance standards to gauge the possibility of data reduction for the respective indicators: Standard 1, Curriculum (seven indicators); Standard 2, Classroom Evaluation and Assessment (eight indicators); and, Standard 3, Instruction (eight indicators). Of note, data reduction is necessary in the current research since the current data sample size of 164 elementary schools ($n = 164$) cannot support the twenty-three collective indicators of the three Academic Performance standards. Therefore, to achieve a more equitable balance of

variables to cases—with the “rule of thumb” being a 15:1 ratio—data reduction is employed. Any variables that loaded together to form a factor were then tested for internal reliability of that factor by measuring the Cronbach’s (1951) alpha—with a Cronbach’s (1951) alpha of 0.7 or higher generally considered the benchmark for claiming the items cluster together well enough to be considered one factor.

Multiple Regression Analysis

Multiple regression analysis offers the researcher a statistical methodology to test the relationship between multiple independent variables and one dependent variable: yielding a single correlation coefficient of the relationship (Vierra, Pollock, & Golez, 1998, pp. 163-164). Multiple regression affords a strong analytical system for the education field by allowing researchers to investigate the “complexity and variety” of relationships: with the opportunity to examine the influence of several independent variables on the dependent variable, the relationships among independent variables, the varying forms of independent variables, and the complexity of relationship forms among independent and dependent variables (Cohen, Cohen, West, & Aiken, 2003, pp. 2-3). In the current study, the researcher performs multiple regression analysis to determine the relationships between the three independent variables (the three *Standards and Indicators for School Improvement* Academic Performance standards of Curriculum, Classroom Evaluation and Assessment, and Instruction) and the single dependent variable (elementary science accountability scores from the Academic Index), while controlling for demographic factors.

Control variables. Control variables, also known as “covariates” and “concomitant variables,” afford researchers the opportunity to “statistically subtract the effects” of influence from a variable, in such analysis as multiple regression (Vogt, 1999, p. 61). According to Huck (2000), control variables are used in regression research to

determine the “‘pure’ relationship between the remaining independent variable(s) and the dependent variable” (p. 591). Huck asserts that control variables allow the researcher to avoid “confounding” the relationship between the independent and dependent variables, if examined directly (p. 591). Thus, in the current research, the control variables, such as Size, Ethnicity, and Appalachian Status, enable the researcher to glean a more accurate appraisal of the impact of curriculum, classroom evaluation/assessment, and instruction on elementary science accountability scores in Kentucky, *ceteris paribus*.

The multiple regression formula. When examining bivariate, linear relationships, researchers employ the simple regression equation, $Y' = a + bX$, whereby Y' represents the predicted score on the dependent variable, a is the constant, b is the regression coefficient, and X is the known value on the independent variable (Huck, 2000, p. 572). However, when using multiple independent variables, as in the current research, Hinkle, Wiersma, and Jurs (2003, p. 461) propose the following multiple regression formula:

$$\hat{Y} = b_1X_1 + b_2X_2 + \dots + b_kX_k + a$$

where a is the regression constant; bs are the regression coefficients (slope of the line) for the respective predictor variables (k); and the given X_k values predict the single criterion variable (\hat{Y}) values. Huck (2000, pp. 581-582) advises that in determining the relative impact of various independent variables, the researcher must analyze something other than an “unstandardized regression equation” and instead a “standardized regression equation” with beta weights can be included. Huck continues by advising “although standardized regression equations are rarely included in research reports, researchers often extract the beta weights from such equations and present the numerical values of these β s” (p. 582).

Multiple regression methodologies. According to Huck (2000, p. 583), “different ‘orders’ in which data on the independent variables can be entered into the analysis”

requires different forms of multiple regression. Huck purports the three most common types of multiple regression are the following: simultaneous multiple regression, stepwise multiple regression, and hierarchical multiple regression (p. 583). In the current study, the researcher will employ simultaneous multiple regression and hierarchical multiple regression.

Simultaneous multiple regression. In simultaneous multiple regression, all data related to the independent variables are considered simultaneously (Huck, 2000, p. 583). Huck describes simultaneous multiple regression with a vegetable soup analogy whereby “all ingredients are thrown into the pot at the same time, stirred, and then cooked together” (p. 583). In calculating multiple regression formulas, the objective of the regression analysis is typically displayed in terms of R , R^2 , or adjusted R^2 (Huck, p. 585). According to Huck, the R of multiple regression is computed in the same way as the r in bivariate regression: “Multiple R is what we would get if we computed Pearson’s r between Y and Y' scores for the individuals who provided scores on the independent and dependent variables” (p. 585). In research, Huck details that researchers are far more likely to report R^2 or the percentage equivalent of R^2 , rather than simply listing the value of R : “the success of the regression analysis is quantified by reporting the proportion or percentage of the variability in the dependent variable that has been accounted for or explained by the study’s independent variables” (p. 585). In simultaneous multiple regression, with all data from all independent variables considered simultaneously, only one R^2 can be computed (p. 585).

Hierarchical multiple regression. To extend the vegetable soup analogy detailed by Huck (2000, pp. 583-584) for simultaneous multiple regression, hierarchical multiple regression is similar to cooking the soup by putting ingredients into the pot in a way as to

emphasize certain ingredients (“flavor and tenderness”): “if we wanted garlic to flavor something else, we’d put it in first even though there’s only a small amount of it required by the recipe. Similarly, we would hold back some of the vegetables (and not put them in with the others) if they are tender to begin with and we want to avoid overcooking them.” According to Huck, hierarchical multiple regression allows independent variables to be entered into analysis in stages (p. 584). Typically, the first independent variables entered are used for control. After accounting for the independent control variables in determining variability in the dependent variable, the remaining variables are incorporated to explain any variance in addition to that already accounted for by the control variables (Huck, 2000, pp. 584-585). In hierarchical multiple regression, unlike simultaneous multiple regression, several R^2 values can be computed: “one for each stage of the analysis wherein individual independent variables or sets of independent variables are added” (Huck, 2000, p. 586). Since several R^2 can be computed in stages with hierarchical multiple regression, researchers determine the change of explained variance as an index, with the notation ΔR^2 (Huck, 2000, pp. 586-587). The notation ΔR^2 uses the Greek letter delta (Δ) for the two-word phrase, “change in” (Huck, 2000, pp. 586-587).

Empirical Research Questions

For the current research, the three empirical research questions that guide this study are repeated here (also, see Figure 1 for the conceptual framework of the relationship among the variables):

To what degree do(es):

1. Demographic Factors relate to Academic Performance (Standard 1, Standard 2, Standard 3) and to the KCCT Science?

2. Academic Performance (Standard 1, Standard 2, Standard 3) relate to the KCCT Science?
3. Academic Performance (Standard 1, Standard 2, Standard 3) relate to the KCCT Science, controlling for the Demographic Factors?

Research Question 1 employs simultaneous multiple regression through four univariate regressions on SISI Standard 1, Standard 2, Standard 3, and the Academic Index (KCCT Science only). Research Question 2 also uses simultaneous multiple regression through one regression equation of Academic Performance (SISI Standards 1-3) on the Academic Index (KCCT Science only). Research Question 3 uses hierarchical multiple regression. In terms of Research Question 3 and the subsequent use of hierarchical multiple regression, only variables that were significant in the prior regressions (Research Question 1 and 2) will be retained for the subsequent hierarchical multiple regression. Table 1 lists the independent and dependent variables, along with the type of multiple regression to be used, by research question. While the elementary science scores from the Academic Index remain the dependent variable for the entire study, due to the formulation of Research Question 1, the Academic Performance standards from the SISI are used as the dependent variable, rather than the elementary science scores from the Academic Index. This distinction of dependent variable helps the researcher eliminate extraneous variables that will not significantly contribute to the elementary science scores of the Academic Index, for the dependent variable on Research Question 3. Additionally, only significant independent variables of demographic factors—and those approaching significance—from Research Question 1 are carried into the hierarchical multiple regression of Research Question 3.

Table 1

Variables and Type of Multiple Regression by Research Question

Empirical Research Questions	Independent Variables		Dependent Variable(s)	Type of Regression
	Control Variables	Academic Performance		
1	DEM V		CURR; CEA; INST	Simultaneous
1	DEM V		AI (ELEM SCI)	Simultaneous
2		CURR; CEA; INST	AI (ELEM SCI)	Simultaneous
3	DEM V	CURR; CEA; INST	AI (ELEM SCI)	Hierarchical

Reliability and Validity

When examining a research study, the reliability and validity of the data prove invaluable in establishing the credibility and impact of the findings. According to Huck (2000, p. 85), “the conclusions drawn and the recommendations made in such studies can be no better than the data on which they are based.” In terms of reliability, Huck summarizes the concept through the word “consistency” (p. 86). While several different techniques exist to evaluate the reliability of an instrument or data, Huck states that each technique attempts to answer the following question: “To what extent can we say that the data are consistent?” (p. 86). Vogt (1999, p. 245) defines reliability as “freedom from measurement (or random) error.” In terms of research, Vogt continues by explaining reliability as “the consistency or stability of a measure or test from one use to the next” (p. 245). According to Evers and Walberg (2004), “reliability is the *sine qua non* of test construction...” (p. 267).

In regards to validity, Huck (2000, p. 100) summarizes the concept with the word “accuracy.” Huck continues by stating that “a researcher’s data are valid to the extent that the results of the measurement process are accurate” and “a measuring instrument is valid to the extent that it measures what it purports to measure” (p. 100). Vogt (1999, p. 301) defines validity as “a term to describe a measurement instrument or test that accurately measures what it is supposed to measure; the extent to which a measure is free of systematic error.” Vogt also states that while reliability is a necessary component of validity, the converse is not true (p. 301).

In the current study, the researcher employed psychometric analyses such as factor analysis as it pertains to reliability and validity—specifically, construct validity—of the data. Huck (2000, p. 104) describes construct validity as a measure to determine “how

much of a personality or psychological construct is possessed by the examinees to whom the instrument is measured.” Factor analysis has proven a valid tool for examining the relationships among a number of variables, when *a priori* consideration is given among the variables (Floyd & Widaman, 1995; Costello & Osborne, 2005). Costello and Osborne (2005, p. 7) state that “optimal results (i.e., results that will generalize to other samples and that reflect the nature of the population) will be achieved by use of a true factor analysis extraction method (we prefer maximum likelihood)...” According to Todd (2010), overall validity of the study, as the current study comes under the larger umbrella of research on the *Standards and Indicators for School Improvement (SISI)* (see B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010), relates to design issues as outlined by Cook and Campbell (1979) and, therefore, would be supported by the accountability model of school improvement for Kentucky.

Regarding threats to validity and reliability, several issues exist within the current research that should be addressed. First, the accountability measure used in this study, the Kentucky Accountability Index, from which the Academic Index, and the subsequent dependent variable of elementary science scores, was extracted, has been examined extensively since its inception with the passage of the Kentucky Education Reform Act (KERA) in 1990. After considerable scrutiny detailing the ineffective and inefficient nature of the first iteration of KERA accountability with the Kentucky Instructional Results Information System (KIRIS), the Kentucky Department of Education (KDE) instituted the Commonwealth Accountability Testing System (CATS) as the alternative measure to remedy issues with reliability and validity that marked KIRIS (Evers & Walberg, 2004).

While CATS underwent examination regarding reliability and validity issues, the

Scholastic Audit process did not incur the same critical analysis for reliability and validity as the aforementioned (Todd, 2010). Koger and Thacker (2004) cited the *Standards and Indicators for School Improvement* (SISI) as the “primary assessment and evaluation instrument for the reviews” (p. 3). Yet, Koger and Thacker determined that the data in the research, results from the Scholastic Audits, had not been verified: “while there were no reliability measures taken to assess inter-rater reliability, all teams did use a standardized instrument and received training on conducting the audits/reviews” (p. 3; p. 80). However, Koger and Thacker stated the type of training and professional development of Scholastic Audit teams differed from year to year (p. 80). In terms of inter-rater reliability for Scholastic Audits, Lyons and Barnett (2011) also cited a weakness in the training: “while the training of audit team members involves multiple days of instruction, the lack of activities to ensure inter-rater reliability weakens comparison between schools audited by different teams” (p. 11).

In addition to the work of Koger and Thacker (2004) in examining Scholastic Audit data, KDE also engaged in an analysis of Scholastic Audit data to guide the use of the SISI in administering Scholastic Audits (Lyons & Barnett, 2011). Upon the administration of Scholastic Audits on schools failing to meet improvement goals, and several successful schools, KDE identified a variety of factors that differentiated the two types of schools. Since KDE believed these factors to be “valuable” for informing future school improvement efforts, KDE subsequently labeled these indicators as “leverage points” and later as “variance points,” since these points “represented variance in audit results between the two groups of schools” (Lyons & Barnett, 2011, p. 1). Lyons and Barnett (2011) explored the variance point concept further in Kentucky schools by differentiating the impact of the KDE identified variance points in each school level—elementary, middle,

and high school. In the analysis of variance points, Lyons and Barnett cited the unclear methodology of KDE in identifying variance points. KDE reported new variance points “after each cycle of audits, with results ranging widely in terms of the number and emphasis of the indicators that were identified as variance points” (Lyons & Barnett, 2011, pp. 1-2). According to Lyons & Barnett (2011, p. 2), citing an investigation conducted by the Appalachian Educational Laboratory (AEL) in 2002, elementary school staff had “greater confidence in, and agreement with, SISI recommendations than middle or high school personnel.” This assertion by Lyons should be noted since the current research pertains to the SISI work during Scholastic Audits administered in elementary schools.

In terms of validity and psychometric quality, when contrasted to the dearth of studies completed by the Kentucky Department of Education (KDE) regarding the Scholastic Audits, the set of completed dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010), conducted under the guidance of Drs. Miller and Smith, provide considerable and rigorous evidence for both construct validity and external criterion validity in the use of the SISI through the Scholastic Audits, with concomitant use of Cronbach’s (1951) alpha for internal scale reliability. Each of the nine SISI standards were further analyzed through factor analysis, specifically at the elementary level (McKinney, Miller, & Smith, 2007; Smith, B. C. Ennis, Saravia, Miller, & Wagner, 2008; Smith, Harvey, Hammock, & Miller, 2009; Todd, 2010). According to Todd (2010), the results of each of these factor analyses confirmed the unified nature of the set of indicators used to describe each of the nine SISI standards. Subsequent multiple regressions of these factors have confirmed the efficacy of these standards, in regards to the Academic Index, as well (Todd, 2010).

To date, no current studies have examined the impact of the first three standards of

the SISI (Standard 1: Curriculum; Standard 2: Classroom Evaluation and Assessment; and, Standard 3: Instruction), also referred to as the Academic Performance standards, on the specific component of elementary science scores from the Academic Index. Smith and Miller (2013) did conduct further analysis on the Scholastic Audit database for an invited presentation on rural schools. This work examined the three larger domains of variables (Academic Performance, Standards 1-3; Learning Environment, Standards 4-6; Efficiency, Standards 7-9) and confirmed the validity of the nine standards as well as their combined effect on the Academic Index.

Ethical Issues

Regarding ethical issues, social science research focuses on maintaining the highest ethical standards through the following: maintaining the confidentiality of subjects, minimizing or eliminating the potential psychological and physical harm to subjects, and providing informed consent of the research purpose to subjects (Vierra, Pollock, & Golez, 1998). The current study was presented to the Institutional Review Boards of the University of Louisville and Western Kentucky University to assure the aforementioned ethical considerations (see Appendix B).

In addition, the current research was extracted from the repository of Scholastic Audit data collected by the audit teams at the Kentucky Department of Education. As secondary data, it is important to note that the researcher did not directly collect the data for this study. The current research poses no risk to subjects as no K-12 educational institution, teachers, or students were identified.

Summary

With an ever-growing emphasis on Science, Technology, Engineering, and Math (STEM) education to meet the exponential rise of technical jobs for the 21st century,

America is looking now, more than ever, to address the demand for college and career ready students with enhanced classroom instruction.

This study is a quantitative analysis of Scholastic Audit secondary data for the relationships of Academic Performance standards (1: Curriculum; 2: Classroom Evaluation and Assessment; and, 3: Instruction) from the *Standards and Indicators for School Improvement* (SISI) and the subsequent impact on elementary science scores, as extracted from the Commonwealth Accountability Testing System (CATS) results by the Kentucky Department of Education (KDE). In this study, the dependent variable is the school-level elementary science scores from the Academic Index, constituting 90.25% of the larger Accountability Index, for Kentucky. The remaining portion of the Accountability Index includes the Norm-Referenced Test (NRT) accounting for 5% and the Nonacademic Index (retention and attendance) accounting for the remaining 4.75% (KDE, 2006a, Appendix D).

Scholastic Audit data are collected by trained audit teams who use the SISI as the primary instrument for evaluation and assessment (Koger & Thacker, 2004, p. 3). The SISI is composed of nine standards, further grouped into three domains: Academic Performance standards (Standard 1, Curriculum; Standard 2, Classroom Evaluation and Assessment; and, Standard 3, Instruction), Learning Environment standards (Standard 4, School Culture; Standard 5, Student, Family and Community Support; and, Standard 6: Professional Development, Professional Growth and Evaluation), and Efficiency standards (Standard 7, Leadership; Standard 8, Organizational Structure and Resources; and, Standard 9, Comprehensive and Effective Planning).

Again, the current research only examines the Academic Performance standards of the SISI and the subsequent impact on the dependent variable of elementary science scores,

as extracted from the Academic Index of CATS scores in Kentucky. For clarity, demographic control variables are also used to approximate more accurately the relative impact of the Academic Performance standards on elementary science scores. The current research maintains the highest level of ethical considerations through the use of a secondary database, with appropriate approval from the Institutional Review Boards of the University of Louisville and Western Kentucky University. While other dissertations have analyzed the impact of SISI standards on the Academic Index at varying school levels in Kentucky (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010), this research is the only study to examine the influence of the SISI Academic Performance standards on elementary science scores.

CHAPTER IV

RESULTS

Introduction

The purpose of this study is to investigate the following: (a) any relationships between demographic factors to Academic Performance and to science achievement in Kentucky elementary schools; (b) any relationships between Academic Performance to science achievement in Kentucky elementary schools; and (c) any relationships between Academic Performance to science achievement in Kentucky elementary schools, when controlling for demographic factors. The central research question driving this study is the following: To what extent is Academic Performance (as defined by Standards 1-3 on the *Standards and Indicators for School Improvement*) related to elementary science scores on the Kentucky Core Content Test (KCCT)?

The researcher utilized secondary data from a Scholastic Audit database maintained by the Kentucky Department of Education (KDE), and demographic data school scores from the Kentucky Performance Report (KPR), to investigate the posed questions. The empirical research questions are graphically represented through Figure 1, the conceptual framework (Chapter I, p. 17), and serve to guide the study. The Scholastic Audits were a school reform process conducted at low-performing schools, and a sampling of voluntary audits at high-performing schools, to help identify areas of growth for these schools, to assist them in reaching the state goal of 100 on the Accountability Index by 2014.

Scholastic Audits were grounded in the *Standards and Indicators for School Improvement*

(SISI), a whole-school reform document constituted of nine standards across 88 indicators. The SISI was sub-divided into three domains of school improvement, each with three standards: (a) Academic Performance: Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction); (b) Learning Environment: Standard 4 (School Culture), Standard 5 (Student, Family and Community Support), and Standard 6 (Professional Growth, Development and Evaluation); and (c) Efficiency: Standard 7 (Leadership), Standard 8 (Organizational Structure and Resources), and Standard 9 (Comprehensive and Effective Planning). A complete list of the SISI standards and indicators are listed in Appendix A. As previously noted, for the current research, Academic Performance refers only to the first three standards of the SISI, and not as an indication of student achievement or school performance on testing measures.

The study uses 181 Kentucky elementary schools ($N = 181$) that were audited between 2001-2005 as the sample. Furthermore, 17 repeat audit incidences were removed to ensure the independence of study design, adjusting the study sample ($n = 164$)—a unique feature of this study compared with its peers (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010). For the present research, the school is the unit of analysis. Elementary school science scores from the Academic Index of the Commonwealth Accountability Testing System (CATS) served as the dependent variable. The Academic Performance standards from the SISI—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—served as the target independent variables. Demographic factors were used as control measures.

The researcher employed a correlational design for the current study. Descriptive statistics were used for investigating the independent variables, such as the demographic control variables, which are described in further detail in the following section. For the

Academic Performance standards from the SISI, descriptive statistics are presented in the section on **Psychometric Analysis**. Factor analysis and Cronbach's (1951) alpha were used to find if any reduction in the variables was necessary for variable to case ratio and to measure the internal reliability of the variables, respectively. More specifically, factor analysis was conducted to determine if any or all of the indicators for the respective SISI Academic Performance standards could be combined into an index—data reduction. Furthermore, using the criterion of a Cronbach's (1951) alpha of 0.7 or higher being statistically significant (Vogt, 1999, p. 64), the researcher also tested the internal reliability of the proposed index. Multiple regression was utilized for the study to determine the relationships between the three independent variables, the SISI Academic Performance standards, and the dependent variable, elementary science scores from the Academic Index, while controlling for any demographic factors. Additionally, a two-step hierarchical multiple regression was specifically used by first entering the demographic factors and, after accounting for the variance on the Academic Index by these control variables, then entering the SISI Academic Performance standards to approximate more accurately the net influence of these three standards on the elementary science accountability scores.

Descriptive Statistics

In the current section, the researcher provides a detailed list of the descriptive statistics for the demographic control variables. As previously mentioned, the *Standards and Indicators for School Improvement* (SISI) Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—are described statistically in the following section on **Psychometric Analysis**. The researcher employed secondary data in the study, as collected during Scholastic Audit and Reviews at Kentucky elementary schools by Kentucky Department of Education

(KDE) trained audit teams. In terms of the methodology for the Scholastic Audits, data were collected with the SISI as the central assessment and evaluation tool. Additionally, Scholastic Audit teams specifically collected the following types of evidence, in alignment with the indicators: surveys, portfolios, team experiences, interviews, and anecdotal observations (KDE, 2003). The Audit teams then collectively weighed all this evidence at the conclusion of the external audit, coming to consensus on a single score on a 4-point scale (with 4 = high) for each of the 88 indicators across the nine standards. As noted, only the scores from the indicators for Standards 1-3 are examined in this study.

For the current study, the researcher analyzed only fourth grade students in terms of the Academic Index component, since fourth grade represented the only grade in elementary school where science was assessed. Overall, Scholastic Audits were conducted at 367 schools in Kentucky, representing the total pool of schools receiving an audit—including elementary schools, middle schools, high schools, and two P-12 configurations (Todd, 2010). Since only elementary schools were used for the current study, the overall sample was 181 elementary schools ($N = 181$), with a final adjusted sample ($n = 164$) for the elementary schools once all repeat audits were removed. A repeat Scholastic Audit was conducted at any school that failed to progress from the lowest area of academic assistance during the previous year. During the five rounds of Scholastic Audits between 2000 and 2005, Scholastic Audit teams conducted 56 audits during 2000-2001, 17 audits in 2001-2002, 76 audits during 2002-2003, 5 audits in 2003-2004, and 10 audits during the 2004-2005 round ($n = 164$).

The Kentucky Performance Report (KPR) supplied the data used for the demographic control variables. As noted previously in the current study, sometimes educational researchers attempt to “fit” secondary data around the empirical research

questions—framing the questions (Sautter, 2014, p. 28). Of note, the KDE database in the present study avoids this mistake by storing data that accurately matches the type of data needed for the empirical research questions—namely, Scholastic Audit evidence.

Additionally, no missing or incomplete data were found since Scholastic Audit and Review teams were complete in the collection process. For the dependent variable, elementary science scores (grade four) were extracted from the larger Academic Index that is comprised of all the content areas.

Independent Variables

The variables in the current study were bifurcated into two types of independent variables, as observed in the conceptual model (see Figure 1): (a) control variables, which are the demographics and (b) alterable variables, which are the Academic Performance standards from the SISI—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction (Bloom, 1980).

Demographic controls. Based on the analysis of control variables conducted in prior studies using the *Standards and Indicators for School Improvement (SISI)*, the following demographic controls were used in the study: Size (SIZE), Ethnicity (%WHITE), Appalachian Status (APP), Percentage Free/Reduced Lunch (%FRED), Percentage Non-Disabled Students (%NDAB), Percentage LEP Students (%LEP), Percentage Migrant Students (%MIG), Percentage Gifted/Talented Students (%GFT), Independent/County School District (CTY), Year of Audit (YEAR), Percentage Title I Students (%TITLE1), Percentage Female Students (%FEM), and Percentage Extended School Service Students (%ESS). These 13 demographic variables were included in the current study since these variables are frequently used in other studies that deal with school outcomes. Additionally, these demographic variables align with the total set of control

variables used in the related dissertations that examine the Academic Index and SISI at the elementary (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008) and secondary level (Todd, 2010).

Descriptive statistics for the demographic control variables are presented in Table 2. School size is smaller than might be expected because it was measured through the total number of students in grade 4, the level whereby elementary science was assessed. Of note, the Percentage Free/Reduced Lunch (%FRED) of 61.20% for the 164 audited elementary schools (excluding repeat audits) is significantly above the state average of roughly 50% (Miller, Smith, & L. S. Ennis, 2006). Also, more than 90% of the elementary schools audited are County schools, in contrast to their Independent school district counterparts. Furthermore, 40% of the schools in the sample are identified as Appalachian; nearly one-third of all Kentucky students are considered Appalachian (Smith, 2005, 2006).

Table 2

Descriptive Statistics for Demographic Controls (N = 164)

Measure	<i>M</i>	<i>SD</i>	Minimum	Maximum	Range
SIZE	60.08	--	6.00	145.00	139.00
% WHITE	84.69	--	13.64	100.00	86.36
APP	.40	--	.00	1.00	1.00
%FRED	61.20	22.02	5.15	100.00	94.85
%NDAB	85.73	8.44	26.55	100.00	73.45
%LEP	.66	2.52	.00	22.73	22.73
%MIG	2.29	5.95	.00	50.00	50.00
%GFT	13.78	11.29	.00	50.00	50.00
CTY	.91	--	.00	1.00	1.00
YEAR	2001.37	--	2000.00	2004.00	4.00
%TITLE1	84.04	34.88	.00	100.00	100.00
%FEM	47.59	7.55	21.05	68.42	47.37
%ESS	36.04	22.52	.00	100.00	100.00

While other studies in the current set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) listed the descriptive statistics on the combined sample of schools (with some schools in the sample conducting mandated Audits as Level 3 Assistance Schools and some voluntary), this is the only study to date to provide separate descriptive statistics for both voluntary and non-voluntary schools (see Table 3 and Table 4). Of note, the Free/Reduced Lunch Percentage is significantly higher

for schools required to conduct a Scholastic Audit as opposed to schools who volunteered to conduct a Scholastic Audit, 73.47% to 57.99%, respectively. However, both percentages for Free/Reduced Lunch were above the state average of approximately 50% (Miller, Smith, & Ennis, 2006). Additionally, the percentage of students receiving Title I assistance, another socioeconomic measure, was significantly higher in Forced Scholastic Audit Schools as opposed to Voluntary Scholastic Audit Schools, 97.36% to 80.56%, respectively.

Table 3

Descriptive Statistics for Demographic Controls on Voluntary Scholastic Audit Schools (N = 130)

Measure	<i>M</i>	<i>SD</i>	Minimum	Maximum	Range
SIZE	59.65	--	6.00	145.00	139.00
% WHITE	87.30	--	13.64	100.00	86.36
APP	.42	--	.00	1.00	1.00
% FRED	57.99	22.35	5.15	100.00	94.85
% NDAB	86.15	7.11	60.71	100.00	39.29
% LEP	.66	2.71	.00	22.73	22.73
% MIG	2.35	5.39	.00	50.00	50.00
% GFT	15.45	11.35	.00	50.00	50.00
CTY	.93	--	.00	1.00	1.00
YEAR	2001.52	--	2000.00	2004.00	4.00
% TITLE1	80.56	37.65	.00	100.00	100.00
% FEM	47.87	7.52	21.05	68.42	47.37
% ESS	36.83	22.47	.00	100.00	100.00

Table 4

Descriptive Statistics for Demographic Controls on Forced Scholastic Audit Schools (N =34)

Measure	<i>M</i>	<i>SD</i>	Minimum	Maximum	Range
SIZE	61.71	--	21.00	130.00	109.00
% WHITE	74.75	--	33.33	100.00	66.67
APP	.32	--	.00	1.00	1.00
%FRED	73.47	15.75	42.03	95.24	53.21
%NDAB	84.14	12.31	26.55	98.51	71.96
%LEP	.68	1.59	.00	6.19	6.19
%MIG	2.04	7.81	.00	45.61	45.61
%GFT	7.37	8.56	.00	32.84	32.84
CTY	.82	--	.00	1.00	1.00
YEAR	2000.76	--	2000.00	2002.00	2.00
%TITLE1	97.36	15.41	10.14	100.00	89.86
%FEM	46.50	7.69	28.57	66.67	38.10
%ESS	33.01	22.80	.00	100.00	100.00

Dependent Variable

In this section, the researcher provides descriptive statistics for the dependent variable, the Academic Index—specifically, elementary science scores from the Academic Index. The Academic Index constituted 90.25% of the overall Accountability Index. The Academic Index measured school performance on the assessments administered by the

Kentucky Department of Education (KDE) through the Kentucky Core Content Tests (KCCT) and under the Commonwealth Accountability Testing System (CATS). By 2014, each school needed to achieve the goal of proficiency (100). For the current study, the mean and standard deviation for the overall sample on the elementary science portion of the Academic Index ($n = 164$) are 69.11 and 14.77, respectively. For schools forced to conduct a Scholastic Audit, the mean and standard deviation for the sample on the elementary science assessment ($n = 34$) are 52.89 and 8.09, respectively. And, for the elementary schools who were not forced to conduct a Scholastic Audit, but rather volunteered, the mean and standard deviation of the sample for elementary science scores ($n = 130$) are 73.36 and 13.09, respectively.

Psychometric Analysis

Within the *Standards and Indicators for School Improvement* (SISI), the nine standards and respective indicators served as a guide for whole school reform and improvement. Outside the dissertations within this research program (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010), almost no research has been conducted, even by the KDE, on this set of standards and indicators to confirm the efficacy of the standards—especially in terms of data reduction. (Smith and Miller, 2013, examined all nine standards but in the context of a new rural (place) measure.) Further, the sample for this study differs from the other related dissertations because of the omission of the second Scholastic Audit for the 17 schools that were examined twice. These conditions necessitated the computation of Factor Analysis and Cronbach's (1951) alpha by the researcher to assess the possibility of data reduction and to the internal reliability of any subsequent factors, respectively, for this unique data set.

Factor Analysis

To gauge the extent to which items represent an underlying construct (factor) among a set of variables, the current study uses factor analysis as a means of data reduction. Nunnally and Bernstein (1994) noted that “relating variables to underlying groupings, testing for groupings, or discovering groupings can involve anything from a broad, essentially atheoretical, data-driven search to testing a highly developed theory” (p. 451). According to the researchers, factor analysis was an appropriate methodology for these diverse needs. In a discussion of the best practices in exploratory factor analysis, Costello and Osborne (2005) cited factor analysis as a common statistical method in the social sciences (p. 1). The following paragraphs discuss factor analysis through three lenses: as a means of factor extraction, retention, and rotation, plus the use of factor scores in subsequent analyses.

While principal components analysis (PCA) was noted as the default extraction method in such statistical software programs as SPSS, Costello and Osborne (2005) warned against PCA: “PCA is not a true method of factor analysis and there is disagreement among statistical theorists about when it should be used, if at all” (pp. 1-2). Costello and Osborne proffered factor analysis as the preferred method versus PCA, since PCA is only a data reduction method (p. 2). Contrastingly, factor analysis helps researchers determine if any latent variables are causing the manifest variables to covary. Additionally, Costello and Osborne stated, “since factor analysis only analyzes shared variance, factor analysis should yield the same solution (all other things being equal) while also avoiding the inflation of estimates of variance accounted for [unlike PCA]” (p. 2). The current research uses maximum likelihood (ML) exploratory factor analysis. Nunnally and Bernstein (1994) cited ML as a popular approach for using “sample data to predict the results in a population” (p. 469). ML differs from PCA in that ML stresses “statistical

inference rather than assuming an indefinitely large sample” (p. 469).

Following the extraction of factors, Costello and Osborne (2005) found the need for determining the methodology for which factors to retain—cautioning that “both overextraction and underextraction of factors retained for rotation can have deleterious effects on the results” (p. 2). While most statistical software programs retain factors with eigenvalues greater than 1.0, Costello and Osborne cited evidence that only using this criterion for factor retention is one of the least accurate methods. Subsequently, the researchers discovered that the scree test is the best method for determining which factors to retain (p. 3). The use of a scree test involves examining a graph of eigenvalues and determining where the data curve flattens out. Thus, the number of factors above this “break” in the data indicate the number of factors to retain. Thus, the current study employs scree tests as a method for ascertaining factor retention.

Of note, and as mentioned in Chapter III, data reduction is a requisite step in the current research to achieve a certain ratio of variables to cases—with 15 variables to one case being considered ideal. Since the current data sample size of 164 schools ($n = 164$) cannot support the set of twenty-three indicators for the three Academic Performance standards—Standard 1, Curriculum, with seven indicators; Standard 2, Classroom Evaluation/Assessment, with eight indicators; and Standard 3, Instruction, with eight indicators—data reduction became necessary to achieve a more equitable balance—in line with a 15:1 ratio.

After the extraction and retention method have been confirmed for the research, Costello and Osborne (2005) suggest that the researcher decide upon a rotation method. Factor rotation affords the research enhanced clarity with the data. However, rotation does not “improve the basic aspects of the analysis” (p. 3). With varimax and other orthogonal

rotations yielding uncorrelated factors, Costello and Osborne recommend an oblique rotation method for its ability to produce correlated factors. While orthogonal rotations might deliver more easily understandable results, the correlated nature of social sciences itself lends itself to an oblique rotational method: “using orthogonal rotation results in a loss of valuable information if the factors are correlated, and oblique rotation should theoretically render a more accurate, and perhaps more reproducible, solution” (p. 3). Subsequently, the current study uses an oblique rotation for clarifying data.

Psychometric analyses of the indicators for each of the three standards determined that each standard loaded onto a single respective factor. Therefore, factor score weighted scales constructed from factor score coefficients were used to represent Standards 1, 2, and 3 in subsequent regression analyses to investigate the three empirical research questions. In terms of regression analyses for the current study, Research Questions 1 and 2 employ simultaneous multiple regression, while Research Question 3 uses hierarchical multiple regression as described in Figure 1 and Table 1.

Academic Performance

In the present study, the three standards that constitute the Academic Performance domain of the *Standards and Indicators for School Performance (SISI)*—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—were each respectively assessed through Factor Analysis to determine the extent to which data reduction were possible. SISI Standard 1, Curriculum, is composed of seven indicators. Conducting a factor analysis on these seven indicators, a single factor emerged that accounted for 57.43% of the total variance among the indicators. The total variance explained for Standard 1, Curriculum, through an extraction method of maximum likelihood, is listed in Table 5.

Table 5

Total Variance Explained for Standard 1, Curriculum

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.020	57.429	57.429	3.532	50.464	50.464
2	.824	11.767	69.196			
3	.577	8.240	77.436			
4	.531	7.583	85.019			
5	.373	5.332	90.351			
6	.364	5.203	95.554			
7	.311	4.446	100.000			

Note. Extraction Method: Maximum Likelihood.

In Figure 2, the researcher included the scree plot to represent graphically the appropriate number of factors, based on eigenvalues, above the break in the line.

Subsequently, the following scree plot confirmed a one factor solution, in accordance with the extraction method.

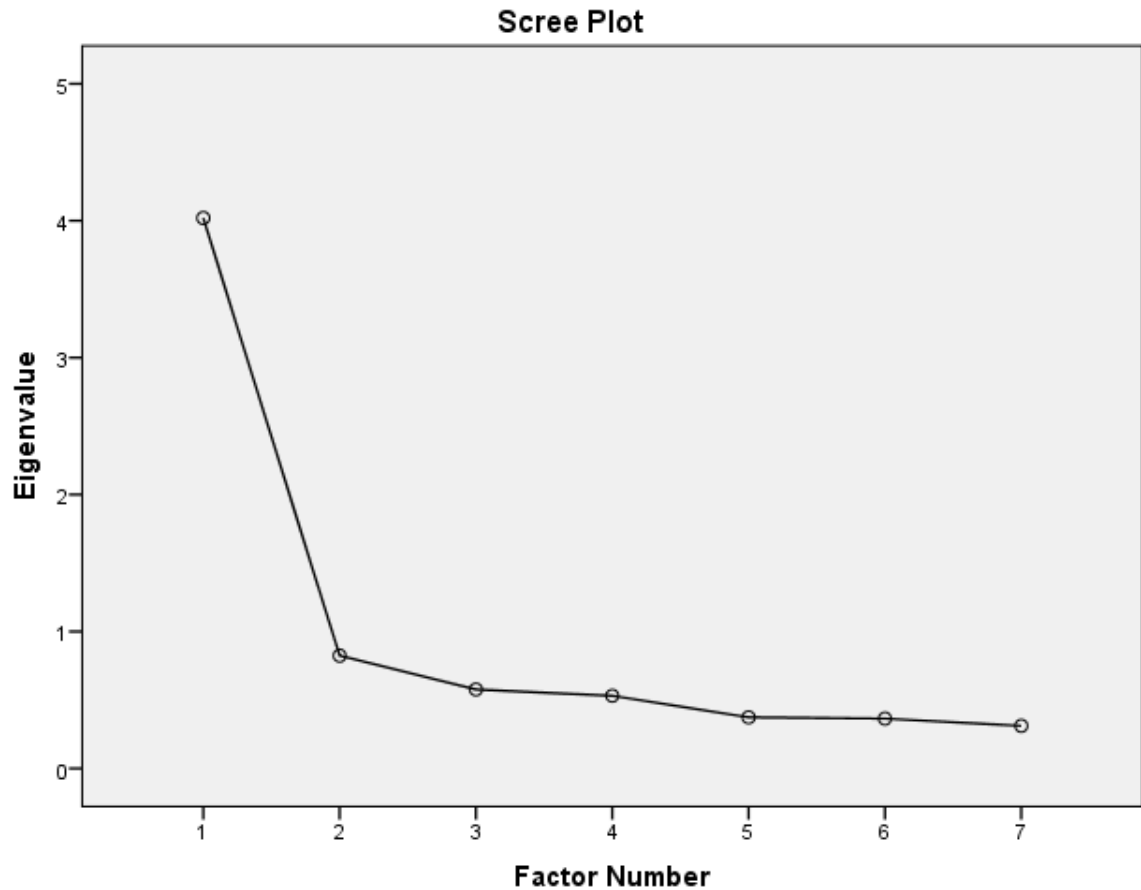


Figure 2. Scree plot for Standard 1, Curriculum.

In Table 6, the seven indicators for Standard 1, Curriculum, and the respective factor loadings are listed. According to Field (2009), the underlying assumption in factor analysis is that “factors represent real-world dimensions, the nature of which must be *guessed at* by inspecting which variables have high loads on the same factor” (p. 633). Examining the loadings from the factor analysis of the one component solution, four of the seven indicators cluster together tightly (1.1.g, 1.1.c, 1.1.f, and 1.1.b), ranging from .770 to .728. While the other indicators do not cluster together as tightly, all indicators add to the overall one factor solution for Curriculum.

Table 6

Factor Loadings for Standard 1, Curriculum

Indicators ^a	Loadings
1.1.g The curriculum provides access to a common academic core for all students.	.770
1.1.c The district initiates and facilitates discussions between schools in the district in order to eliminate unnecessary overlaps and close gaps.	.764
1.1.f There is in place a systematic process for monitoring, evaluating and reviewing the curriculum.	.751
1.1.b The district initiates and facilitates discussions among schools regarding curriculum standards to ensure they are clearly articulated across all levels (P-12).	.728
1.1.d There is evidence of vertical communication with an intentional focus on key curriculum transition points within grade configurations (e.g., from primary to middle and middle to high).	.672
1.1.a There is evidence that the curriculum is aligned with the <i>Academic Expectations, Core Content for Assessment, Transformations</i> and the <i>Program of Studies</i> .	.648
1.1.e The school curriculum provides specific links to continuing education, life and career options.	.626

^aKDE (2008, pp. 3-9).

The one factor data reduction of Standard 1, Curriculum, yielded a Cronbach's coefficient alpha of .874, which indicates high internal reliability—confirming the factor analysis results. A listing of the psychometric properties—including mean, standard deviation, range, and the weighted factor score index—of SISI Standard 1, Curriculum, is included below in Table 7.

Table 7

Internal Reliability and Item Characteristics for Standard 1, Curriculum (N = 164)

Indicator	<i>M</i>	<i>SD</i>	Range	$\alpha - d^a$
1.1.a	2.78	1.79	7	.862
1.1.b	2.37	1.59	7	.856
1.1.c	2.37	1.72	7	.851
1.1.d	1.85	1.54	7	.860
1.1.e	2.48	1.85	7	.865
1.1.f	1.84	1.62	7	.854
1.1.g	2.20	1.75	7	.847
Standard	2.27	1.69	7	.874 ^b

^a $\alpha - d$ = alpha with item deleted.

^bCronbach's coefficient alpha for the entire scale is given for the Standard.

Standard 2, Classroom Evaluation/Assessment, is the second standard of the Academic Performance domain and is composed of eight indicators. Like Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment) also loaded onto one single factor. This single factor explained 56.11% of the total variance among the indicators, as seen in Table 8. The Standard 2 indicators loaded as a single factor. The subsequent inclusion of a scree plot, factor loadings, and Cronbach's coefficient alpha confirms this decision.

Table 8

Total Variance Explained for Standard 2, Classroom Evaluation/Assessment

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.489	56.109	56.109	4.023	50.291	50.291
2	.891	11.136	67.245			
3	.588	7.345	74.591			
4	.559	6.990	81.580			
5	.486	6.069	87.649			
6	.386	4.829	92.478			
7	.315	3.934	96.412			
8	.287	3.588	100.000			

Note. Extraction Method: Maximum Likelihood.

In Figure 3, the scree plot for Standard 2, Classroom Evaluation/Assessment, is included to confirm the one factor solution, based on eigenvalues—as the break above the line indicates one factor. The following scree plot aligns with the extraction method, as both confirm a single component solution.

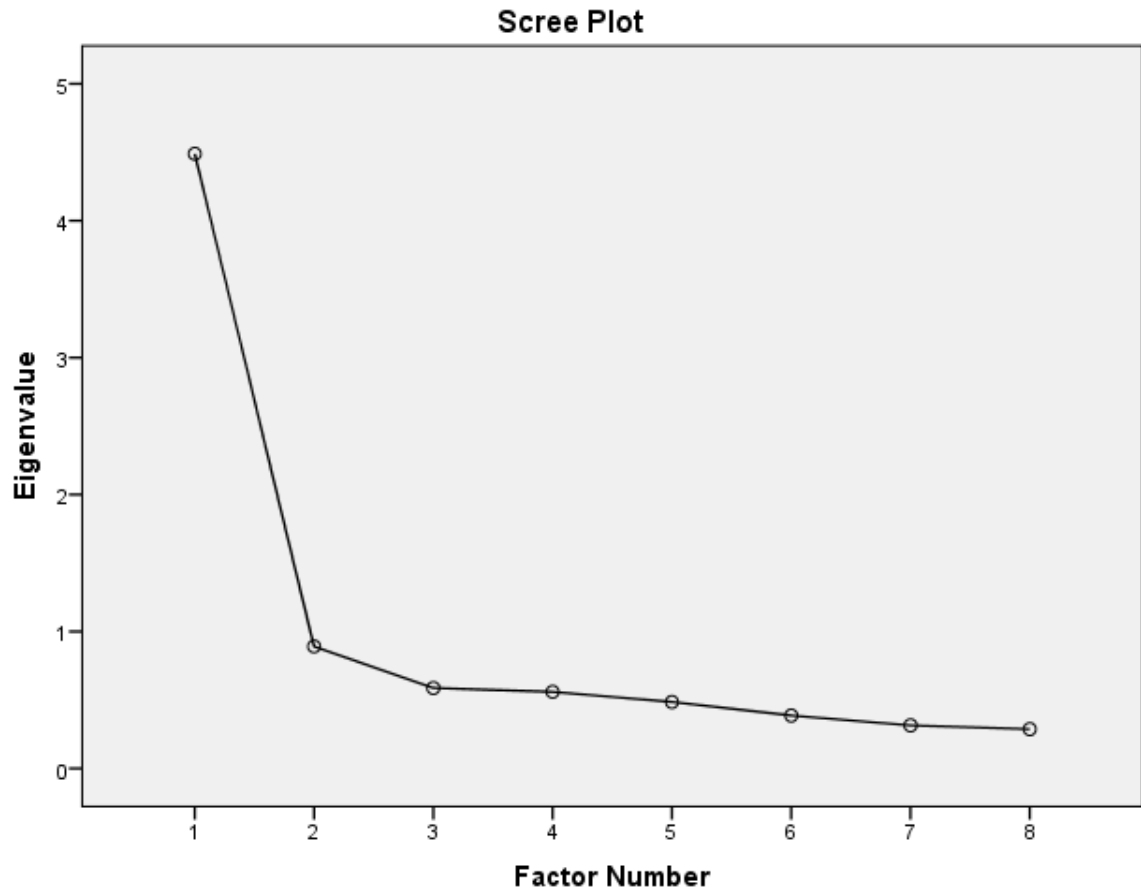


Figure 3. Scree plot for Standard 2, Classroom Evaluation/Assessment.

Table 9 includes the eight indicators for Standard 2, Classroom Evaluation/Assessment, and the subsequent factor loadings. Five of the eight indicators form a tight cluster (2.1.a, 2.1.c, 2.1.e, 2.1.f, and 2.1.b), ranging from .802 for 2.1.a to .752 for 2.1.b. While the remaining three indicators do not cluster together as cohesively, all indicators contribute to the overall single component solution.

Table 9

Factor Loadings for Standard 2, Classroom Evaluation/Assessment

Indicators ^a	Loadings
2.1.a Classroom assessments of student learning are frequent, rigorous and aligned with Kentucky's core content.	.802
2.1.c Students can articulate the academic expectations in each class and know what is required to be proficient.	.795
2.1.e Multiple assessments are specifically designed to provide meaningful feedback on student learning for instructional purposes.	.779
2.1.f Performance standards are clearly communicated, evident in classrooms and observable in student work.	.778
2.1.b Teachers collaborate in the design of authentic assessment tasks aligned with core content subject matter.	.752
2.1.h Samples of student work are analyzed to inform instruction, revise curriculum and pedagogy, and obtain information on student progress.	.655
2.1.d Test scores are used to identify curriculum gaps.	.563
2.1.g Implementation of the state-required Assessment and Accountability Program is coordinated by school and district leadership.	.473

^aKDE (2008, pp. 10-17).

In terms of internal reliability, an overall Cronbach's coefficient alpha of .879 was calculated for Standard 2, determined as a high degree of reliability. Thus, the Standard 2 indicators loaded as a single factor. Table 10 displays the psychometric properties for internal reliability of Standard 2, Classroom Evaluation/Assessment.

Table 10

*Internal Reliability and Item Characteristics for Standard 2, Classroom**Evaluation/Assessment (N = 164)*

Indicator	<i>M</i>	<i>SD</i>	Range	$\alpha - d^a$
2.1.a	1.84	1.41	7	.855
2.1.b	1.39	1.33	5	.861
2.1.c	1.37	1.48	7	.856
2.1.d	2.20	1.84	7	.877
2.1.e	1.54	1.48	5	.856
2.1.f	1.83	1.45	7	.858
2.1.g	3.67	1.67	7	.883
2.1.h	1.81	1.43	5	.867
Standard	1.96	1.51	6.25	.879 ^b

^a $\alpha - d$ = alpha with item deleted.

^bCronbach's coefficient alpha for the entire scale is given for the Standard.

The final Academic Performance standard from the SISI, Standard 3, Instruction, is composed of eight indicators. After conducting a factor analysis on the eight indicators through an extraction method of maximum likelihood, the researcher determined that the data could be reduced to a single factor, constituting 52.35% of the total variance among the indicators, as listed in Table 11. Additionally, the later inclusion of a scree plot, factor loadings, and Cronbach's coefficient alpha confirms the decisions for a single component solution.

Table 11

Total Variance Explained for Standard 3, Instruction

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.188	52.349	52.349	3.669	45.857	45.857
2	.775	9.685	62.034			
3	.724	9.051	71.085			
4	.613	7.660	78.745			
5	.585	7.310	86.055			
6	.445	5.564	91.619			
7	.378	4.725	96.344			
8	.293	3.656	100.000			

Note. Extraction Method: Maximum Likelihood.

The following scree plot in Figure 4 aligns with the determination for a single factor solution, as the break above the eigenvalue line indicates one factor. The scree plot confirms the previous decision, based on the extraction method of maximum likelihood analysis, to use a single component solution.

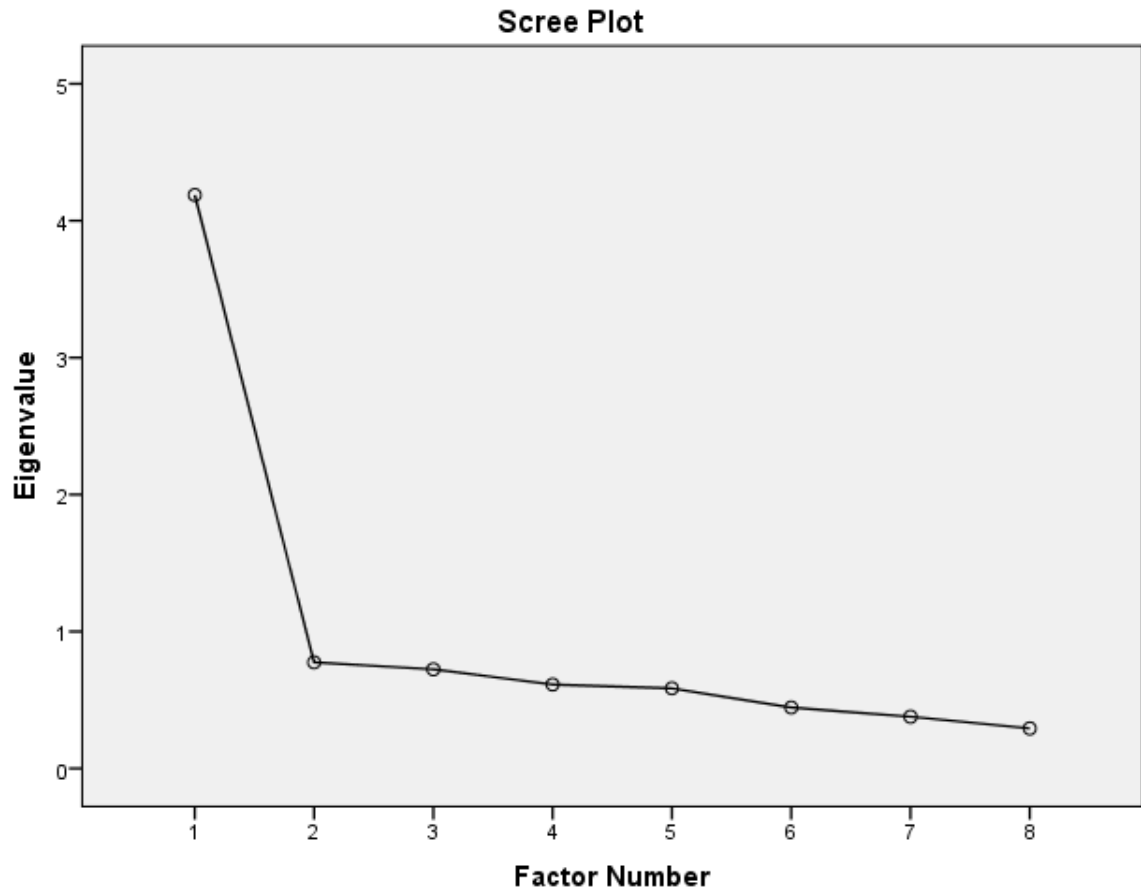


Figure 4. Scree plot for Standard 3, Instruction.

Table 12 lists the factor loadings for SISI Standard 3, Instruction. Upon analysis of the individual loadings of the eight indicators of Standard 3 in the one factor solution, six of the eight indicators cluster together tightly (3.1.c, 3.1.a, 3.1.b, 3.1.d, 3.1.f, and 3.1.h), ranging from .833 for 3.1.c to .625 for 3.1.h. While 3.1.g and 3.1.e do not cluster together as tightly as their counterparts, all indicators contribute to the overall single component solution for Standard 3, Instruction.

Table 12

Factor Loadings for Standard 3, Instruction

Indicators ^a	Loadings
3.1.c Instructional strategies and activities are consistently monitored and aligned with the changing needs of a diverse student population to ensure various learning approaches and learning styles are addressed.	.833
3.1.a There is evidence that effective and varied instructional strategies are used in all classrooms.	.762
3.1.b Instructional strategies and learning activities are aligned with the district, school and state learning goals, and assessment expectations for student learning.	.724
3.1.d Teachers demonstrate the content knowledge necessary to challenge and motivate students to high levels of learning.	.665
3.1.f Instructional resources (textbooks, supplemental reading, technology) are sufficient to effectively deliver the curriculum.	.646
3.1.h There is evidence that homework is frequent and monitored and tied to instructional practice.	.625
3.1.g Teachers examine and discuss student work collaboratively and use this information to inform their practice.	.578
3.1.e There is evidence that teachers incorporate the use of technology in their classrooms.	.534

^aKDE (2008, pp. 18-25).

Additionally, an overall Cronbach's coefficient alpha of .865 for Standard 3, Instruction, confirmed the single factor as having a high degree of internal reliability. The psychometric properties of SISI Standard 3, Instruction, are listed in Table 13.

Table 13

Internal Reliability and Item Characteristics for Standard 3, Instruction (N = 164)

Indicator	<i>M</i>	<i>SD</i>	Range	$\alpha - d^a$
3.1.a	1.73	1.41	7	.842
3.1.b	2.05	1.65	7	.842
3.1.c	1.67	1.36	5	.835
3.1.d	3.15	1.78	7	.848
3.1.e	1.83	1.47	5	.860
3.1.f	3.10	1.76	7	.850
3.1.g	1.61	1.36	7	.855
3.1.h	2.06	1.57	5	.854
Standard	2.15	1.55	6.25	.865 ^b

^a $\alpha - d$ = alpha with item deleted.

^bCronbach's coefficient alpha for the entire scale is given for the Standard.

Representing all dependent, independent, and control variables in the current study, a correlation matrix is presented in Table 14 to capture all possible correlations. The correlation matrix data shows very strong inter-correlations among the three SISI standards in this study—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—with Standards 1 and 2 possessing a correlation of $r = .681$, Standards 1 and 3 reflecting a correlation of $r = .690$, and Standards 2 and 3 demonstrating a correlation of $r = .811$, all at the $p \leq .01$ level. Todd (2010) described the external criterion validity among the standards in the SISI, based on high school audits, as they

possessed high levels of inter-correlations, as “ideal”—with each standard related to others yet still acting independently in the whole school reform model of the SISI (p. 144). That judgment must be moderated slightly for this elementary sample with repeat audits deleted, as the value for Standard 2 with Standard 3 suggests less than ideal independence between these two standards. The correlation among the first three SISI standards further confirms their domain grouping under Academic Performance—those measures that can be addressed in the classroom.

Overall, outside of the inter-correlations among the first three standards of the SISI, most correlations are weak to moderate, with an approximate range of .10 to .70. Each of the three Academic Performance standards purports moderate correlations to the Science Academic Index (ScAI): $r = .390$ for Standard 1 (Curriculum); $r = .526$ for Standard 2 (Classroom Evaluation/Assessment); and $r = .501$ for Standard 3 (Instruction), all at the $p \leq .01$ level. Additionally, several demographic and control variables such as Year of Audit, Percent White, Percent Title I, Percent Gifted, Percentage Free/Reduced Lunch, and Voluntary versus Involuntary status are moderately correlated to the ScAI. Among the aforementioned demographic and control variables, the highest association to the Science scores is the Voluntary versus Involuntary status, with negative correlation of $r = -.564, p \leq .01$. Thus, if a school was forced to conduct a Scholastic Audit (involuntary), the likelihood would increase that this school would score lower on the science component of the Academic Index—aligning with the notion that lower-performing schools which were mandated to perform a Scholastic Audit would subsequently score lower on the Academic Index. Additionally, Percentage Free/Reduced Lunch, a measure of socioeconomic status, also reflects a negative relationship ($r = -.541, p \leq .01$) with ScAI—with inclusion in this demographic factor resulting in lower test scores. Of note, Percent Gifted, Percent Title 1,

and Percentage Free/Reduced Lunch all have significant correlations with almost every other demographic factor. In contrast, the demographic factor of Percent Female has no significant correlations with any of the other demographic variables. Also, Percent Non-Disabled has significant correlations only with Percent Gifted and Percentage Free/Reduced Lunch. Finally, of note is the correlation between Percentage Free/Reduced Lunch and Percent Title I at $r = .653$, as both represent operationalizations of poverty level. In general, the correlations with Percentage Free/Reduced Lunch and other variables are higher than the parallel correlations with Percent Title I, suggesting that the measure of lunch eligibility is the more robust of the two.

Table 14

Correlations for Demographic Factors, Curriculum (Standard 1), Classroom Evaluation & Assessment (Standard 2), Instruction (Standard 3), and Academic Index (N = 164)

	AI	CURR	CEA	INST	YEAR	CTY
AI	--	.390**	.526**	.501**	.512**	.236**
CURR		--	.681**	.690**	.297**	-.017
CEA			--	.811**	.320**	.104
INST				--	.284**	.129
YEAR					--	.100
CTY						--
APP						
SIZE						
%FEM						
%WHITE						
%TITLE1						
%MIG						
%LEP						
%ESS						
%GFT						
%FRED						
%NDAB						
VOL						

(table continues)

Table 14. (continued)

	APP	SIZE	%FEM	% WHITE	%TITLE1	%MIG
AI	-.012	.104	.051	.355**	-.339**	-.023
CURR	-.132	.083	.043	-.005	-.249**	-.011
CEA	-.141	.146	.137	.044	-.308**	-.007
INST	-.205**	.100	.141	.033	-.274**	-.072
YEAR	-.184*	.041	-.061	-.066	-.104	-.054
CTY	.131	.096	.000	.205**	-.066	-.108
APP	--	-.281**	.038	.515**	.255**	-.013
SIZE		--	.034	-.167*	-.277**	-.018
%FEM			--	-.002	.028	.058
% WHITE				--	-.090	.035
%TITLE1					--	.090
%MIG						--
%LEP						
%ESS						
%GFT						
%FRED						
%NDAB						
VOL						

(table continues)

Table 14. (continued)

	%LEP	%ESS	%GFT	%FRED	%NDAB	VOL
AI	.003	.016	.506**	-.541**	.144	-.564**
CURR	.075	-.090	.160*	-.254**	.133	-.281**
CEA	.006	.024	.235**	-.340**	.041	-.277**
INST	.051	.057	.263**	-.358**	.078	-.300**
YEAR	.230**	.046	.383**	-.235**	.080	-.265**
CTY	-.192*	-.040	-.073	-.126	.025	-.151
APP	-.212**	.107	-.021	.393**	.016	-.082
SIZE	-.043	-.246**	.038	-.415**	.066	.029
%FEM	-.006	.004	-.001	.010	.005	-.074
% WHITE	-.484**	-.037	.163*	-.199*	.040	-.254**
%TITLE1	.038	.212**	-.188*	.653**	-.109	.196*
%MIG	-.075	.008	-.124	.044	-.033	-.021
%LEP	--	.109	.055	.158*	-.094	.004
%ESS		--	-.035	.307**	-.019	-.069
%GFT			--	-.390**	.187*	-.291**
%FRED				--	-.213**	.286**
%NDAB					--	-.097
VOL						--

* $p \leq .05$. ** $p \leq .01$.

Research Questions

The current study was guided by three empirical research questions which sought

to communicate relationships among demographic factors and the first three standards of the SISI (Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction) to the elementary science component of the Academic Index. As previously mentioned, psychometric analyses of the respective indicators for each of the three Academic Performance standards yielded a single factor solution whereby factor score weighted scales were used for Standards 1, 2, and 3 in subsequent regression analyses (Riley, 2016).

Multiple regression analysis affords the researcher the opportunity to investigate the relationships between one or more independent variables on a dependent variable, while controlling for one or more control variables (Slavin, 2007). In differentiating between the two multiple regression analyses used in the current study, simultaneous multiple regression (sometimes referred to as “forced entry” regression) occurs when all predictor variables are entered into the regression equation at the same time; in contrast, hierarchical multiple regression occurs when the researcher makes specific decisions about when to enter specific predictors into the regression equation. Additionally, hierarchical multiple regression enables the researcher to enter first the known predictors in the order of importance for predicting the outcome (Field, 2009, p. 212).

For the current study, the researcher employed the statistical software package, SPSS, to calculate the multiple regression models in determining the impact of the first three standards of the SISI, the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—on the elementary science component of the Academic Index, while controlling for demographic control variables.

Research Question 1

To what degree do demographic factors relate to Academic Performance (Standard 1, Standard 2, Standard 3) and to KCCT Science?

The results of the regression analyses from Research Question 1 are detailed in Tables 15-18. Tables 15, 16, and 17 assess the impact of demographic factors on Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction), respectively. Table 18 details the effects for demographic factors on the Science Academic Index.

The results of the regression analysis for the impact of demographic variables on Standard 1 (Curriculum) are listed in Table 15. The independent demographic variables included in this regression model are Year of Audit, County vs. Independent School, Appalachian status, Size of School (4th grade), Percent White (4th grade), Percent Title I (4th grade), Percent Gifted (4th grade), Percentage Free/Reduced Lunch (4th grade), Percent Non-Disabled (4th grade), and Voluntary versus Involuntary status. Per Dr. Douglas Smith, methodologist for the current study, Percent Female (4th grade), Percent Migrant (4th grade), Percent LEP (4th grade), and Percent ESS (4th grade) were removed from the regression model after being deemed not significant through the correlation matrix results (personal communication, April 29, 2016). The regression model yielded a significant relationship between the demographic control variables and Standard 1 (Curriculum), $F(10, 163) = 3.468, p < .001$, with two of the independent variables having a significant impact on Curriculum. Both Year of Audit and Voluntary versus Involuntary status yielded significant effects on Standard 1 (Curriculum), with $\beta = .232, p \leq .01$ and $\beta = -.206, p \leq .05$, respectively. Thus, outside of the Year of Audit and Voluntary versus Involuntary status, demographic factors have little influence on Standard 1 (Curriculum), with the Adjusted $R^2 = .131$ confirming a very small effect size.

Table 15

Regression of Standard 1, Curriculum, on the Demographic Variables (N = 164)

Variable	<i>B</i>	<i>SE B</i>	Beta	<i>t</i>	Sig. <i>t</i>
Constant	-374.999	135.407		-2.769	.006
YEAR	.187	.068	.232	2.771	.006
CTY	-.278	.254	-.086	-1.095	.275
APP	-.041	.210	-.022	-.196	.845
SIZE	.000	.003	.011	.132	.895
% WHITE	-.001	.005	-.029	-.279	.781
% TITLE1	-.004	.003	-.156	-1.601	.112
% GFT	-.004	.007	-.053	-.596	.552
% FRED	-.002	.006	-.046	-.337	.737
% NDAB	.009	.008	.080	1.057	.292
VOL	-.475	.190	-.206	-2.504	.013

The regression results for the effects of demographic independent variables on Standard 2 (Classroom Evaluation/Assessment) are listed in Table 16, $F(10, 163) = 4.261$, $p < .001$. With an adjusted R^2 of .167, very little of the variance of Standard 2 (Classroom Evaluation/Assessment) can be accounted for by demographic variables. Only one demographic variable, Year of Audit, had a significant impact, $\beta = .213$, $p \leq .01$ on Classroom Evaluation/Assessment. However, both Percent Title I (4th grade) and Voluntary versus Involuntary status are close to significance with $p = .077$ and $p = .069$, respectively. Much like Standard 1 (Curriculum), Standard 2 (Classroom

Evaluation/Assessment) is relatively independent of influence by demographic control variables.

Table 16

Regression of Standard 2, Classroom Evaluation/Assessment, on the Demographic Variables (N = 164)

Variable	<i>B</i>	<i>SE B</i>	Beta	<i>t</i>	Sig. <i>t</i>
Constant	-348.809	134.300		-2.597	.010
YEAR	.175	.067	.213	2.604	.010
CTY	.140	.252	.043	.557	.579
APP	-.031	.208	-.016	-.148	.883
SIZE	.001	.003	.043	.515	.608
% WHITE	-.001	.005	-.015	-.145	.885
% TITLE1	-.005	.003	-.170	-1.783	.077
% GFT	.004	.007	.051	.589	.557
% FRED	-.004	.006	-.099	-.741	.460
% NDAB	-.005	.008	-.042	-.568	.571
VOL	-.345	.188	-.147	-1.832	.069

In Table 17, the effects of demographic variables on Standard 3 (Instruction) are presented, $F(10, 163) = 4.417, p < .001$. Again, in this regression model, very little variance can be accounted for on Instruction by demographic variables, with an adjusted R^2 of .173. Voluntary versus Involuntary status is the only independent variable that has a significant impact on Standard 3 (Instruction), with $\beta = -.181, p \leq .05$. No other

demographic variables come close to having a significant impact on Instruction, with the nearest significant independent variable being Year of Audit ($p = .135$). Thus, for all the variance accounted for by Academic Performance standards, demographic variables have very little impact on these three SISI standards.

Table 17

Regression of Standard 3, Instruction, on the Demographic Variables (N = 164)

Variable	<i>B</i>	<i>SE B</i>	Beta	<i>t</i>	Sig. <i>t</i>
Constant	-198.798	132.493		-1.500	.136
YEAR	.100	.066	.123	1.504	.135
CTY	.320	.249	.098	1.287	.200
APP	-.269	.205	-.140	-1.311	.192
SIZE	-.001	.003	-.031	-.370	.712
% WHITE	.000	.005	-.007	-.072	.943
% TITLE1	-.003	.003	-.093	-.982	.328
% GFT	.009	.007	.105	1.201	.231
% FRED	-.005	.006	-.123	-.923	.357
% NDAB	.000	.008	-.003	-.037	.971
VOL	-.420	.186	-.181	-2.261	.025

Finally, in terms of Research Question 1, Table 18 lists the regression results for the significant impact of demographic control variables on the elementary science component of the Academic Index, $F(10, 163) = 28.303, p < .001$. With an adjusted R^2 of .626, demographic variables play a significant role on the elementary science portion of the

Academic Index. Five demographic variables have a significant impact on the Academic Index in this regression model: Year of Audit ($\beta = .314, p \leq .01$), Percent White (4th grade) ($\beta = .187, p \leq .01$), Percent Gifted (4th grade) ($\beta = .175, p \leq .01$), Percentage Free/Reduced Lunch (4th grade) ($\beta = -.257, p \leq .01$), and Voluntary versus Involuntary status ($\beta = -.286, p \leq .01$). County versus Independent status also comes close to having a significant role on the Academic Index with a $p = .062$. Thus, demographic factors account for more than 60% of the variance on the elementary science component of the Academic Index.

Table 18

Regression of the Academic Index (Elementary Science) on the Demographic Variables (N = 164)

Variable	<i>B</i>	<i>SE B</i>	Beta	<i>t</i>	Sig. <i>t</i>
Constant	-7923.759	1396.282		-5.675	< .001
YEAR	3.992	.698	.314	5.722	< .001
CTY	4.934	2.620	.097	1.883	.062
APP	.846	2.161	.028	.392	.696
SIZE	.004	.028	.007	.132	.895
% WHITE	.138	.050	.187	2.736	.007
% TITLE1	-.014	.027	-.033	-.516	.607
% GFT	.228	.077	.175	2.983	.003
% FRED	-.172	.060	-.257	-2.873	.005
% NDAB	-.019	.087	-.011	-.214	.831
VOL	-10.339	1.957	-.286	-5.314	< .001

Research Question 2

To what degree does Academic Performance (Standard 1, Standard 2, Standard 3) relate to KCCT Science?

In Table 19, the regression results are presented for the impact of the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—on the elementary science component of the Academic Index. The ANOVA is a significant model, $F(3, 163) =$

22.141, $p < .001$. With an adjusted R^2 of .280, the three Academic Performance standards explain 28% of the variance in elementary student science scores on the Academic Index. Among the three Academic Performance standards, Standard 2, Classroom Evaluation/Assessment, is the best predictor of overall elementary science scores and is the only significant factor, $\beta = .349$, $p \leq .01$. Additionally, Standard 3, Instruction, is close to being significant with a $p = .073$. Thus, through the results of the regression model, the impact of Classroom Evaluation/Assessment and Instruction both yield a far greater impact on the elementary science component of the Academic Index than their Curriculum counterpart.

Table 19

Regression of the Academic Index (Elementary Science) on the SISI Academic Performance Standards (Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction) (N = 164)

Variable	<i>B</i>	<i>SE B</i>	Beta	<i>t</i>	Sig. <i>t</i>
Constant	69.113	.978		70.638	< .001
Standard 1	.040	1.506	.003	.027	.979
Standard 2	5.417	1.840	.349	2.944	.004
Standard 3	3.396	1.880	.217	1.806	.073

Research Question 3

To what degree does Academic Performance (Standard 1, Standard 2, Standard 3) relate to KCCT Science, controlling for the Demographic Factors?

In Table 20, the researcher included all significant variables, and those variables

approaching significance, from Research Question 1 in a hierarchical multiple regression model to demonstrate the impact of Demographic Factors and the Academic Performance standards (Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction) on the elementary science component of the Academic Index. In this hierarchical regression model, the included demographic variables are Year of Audit, County vs. Independent School, Percent White (4th grade), Percent Gifted (4th grade), Percentage Free/Reduced Lunch (4th grade), and Voluntary versus Involuntary status. County versus Independent School status was included in the hierarchical multiple regression equation since it approached significance. Appalachian status, Size of School (4th grade), Percent Title I (4th grade), Percent Non-Disabled (4th grade), Percent Female (4th grade), Percent Migrant (4th grade), Percent LEP (4th grade), and Percent ESS (4th grade) were removed from previous and current regression models after being deemed not significant (Dr. Douglas Smith, personal communication, April 29, 2016; Dr. Stephen Miller, personal communication, July 12, 2016). The full and reduced model F test calculations were completed using the sum of squares method for all models. During Step 1, only demographic factors were added to the model to determine the impact on the elementary science portion of the Academic Index. In Step 2 of the model, the Academic Performance standards were added to determine whether the model improved. With each of the two steps of the hierarchical multiple regression model, the model subsequently improved in its efficacy (equations not reported for each model). Following Step 2 of the hierarchical regression model, the change in the R^2 value is reported.¹

In Step 1, the ANOVA was significant for the hierarchical multiple regression model, $F(6, 163) = 48.160, p < .001$. Demonstrating a strong impact on the Academic Index, the Adjusted R^2 was .634. The demographic control variables of Year of Audit,

Percent White (4th grade), Percent Gifted (4th grade), Percentage Free/Reduced Lunch (4th grade), and Voluntary versus Involuntary status are all statistically significant at the $p < .01$ level. Additionally, County versus Independent School status was also significant at the $p < .05$ level.

During Step 2 of the hierarchical multiple regression model, the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—were added to determine their combined impact, after controlling for demographic control variables. The ANOVA for this model was also significant, $F(9, 163) = 39.435, p < .001$. Step 2 increased the Adjusted R^2 value from .634 to .680. The demographic control variables of Year of Audit, Percent White (4th grade), Percent Gifted (4th grade), Percentage Free/Reduced Lunch (4th grade), and Voluntary versus Involuntary status all remained statistically significant at the $p < .01$ level. Additionally, Standard 2 (Classroom Evaluation/Assessment) had a significant impact on the Academic Index, with $\beta = .214$ at the $p < .01$ level. Neither Standard 1 (Curriculum) nor Standard 3 (Instruction) were significant variables for the model. Also, County versus Independent School status did not maintain its statistical significance, with $p = .082$. Sixty-eight percent of the total variance is accounted for in the final model.

In terms of the impact on the elementary science component of the Academic Index, the researcher employed standardized beta results to indicate the impact of the independent demographic factor on the elementary Science Academic Index. Within the final hierarchical multiple regression model, the following independent demographic control variables produced a corresponding standardized effect on the elementary science portion of the Academic Index, listed from greatest to least effect: Year of Audit ($\beta =$

.262), Voluntary versus Involuntary Status ($\beta = -.251$), Percent White (4th grade) ($\beta = .213$), Percentage Free/Reduced Lunch (4th grade) ($\beta = -.203$), and Percent Gifted (4th grade) ($\beta = .162$). The positive beta for Percent Gifted and the negative beta for Percentage Free/Reduced Lunch aligned with historical data on the impact of these two demographic factors. In the final regression model, for the three Academic Performance standards, only Standard 2 (Classroom Evaluation/Assessment) had a statistically significant impact, after controlling for demographic factors, $\beta = .214, p < .01$.

Interestingly, the beta value for Classroom Evaluation/Assessment had a greater impact on the Academic Index than all demographic factors, save Year of Audit and Voluntary versus Involuntary status—positive findings for a classroom practice that can outweigh the impact of most demographic factors.

Table 20

Regression of the Academic Index (Elementary Science) on the SISI Academic Performance Standards (Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction), Controlling for Demographic Factors (N = 164)

Variable	B	SE B	Beta	t	Sig. t
Step 1					
Constant	-7816.823	1367.609		-5.716	< .001
YEAR	3.937	.683	.310	5.762	< .001
CTY	5.066	2.554	.099	1.983	.049
% WHITE	.147	.038	.200	3.883	< .001
% GFT	.228	.074	.174	3.077	.002
% FRED	-.178	.036	-.266	-4.991	< .001
VOL	-10.497	1.914	-.289	-5.485	< .001

(table continues)

Table 20. (continued)

Variable	<i>B</i>	<i>SE B</i>	Beta	<i>t</i>	Sig. <i>t</i>
Step 2					
Constant	-6616.699	1318.175		-5.020	< .001
YEAR	3.336	.659	.262	5.065	< .001
CTY	4.281	2.444	.084	1.752	.082
% WHITE	.157	.036	.213	4.394	< .001
% GFT	.212	.070	.162	3.021	.003
% FRED	-.136	.035	-.203	-3.947	< .001
VOL	-9.113	1.833	-.251	-4.971	< .001
Standard 1	-.318	1.037	-.020	-.307	.760
Standard 2	3.318	1.242	.214	2.672	.008
Standard 3	.926	1.290	.059	.718	.474

Note. Adjusted $R^2 = .634$ for Step 1; Δ in Adjusted $R^2 = .046$ for Step 2 ($ps < .001$).

Summary

The purpose of this study was to investigate the effects of the three *Standards and Indicators for School Improvement (SISI)* Academic Performance standards on the elementary school KCCT science portion of the Academic Index, while controlling for demographic factors. The current study uses a secondary database of Scholastic Audit data maintained by the KDE. The Scholastic Audit was a means to actualize the whole school reform measures of the SISI document. The data were collected between 2001 and 2005 by KDE-trained teams of educators. While there were 760 elementary school in Kentucky

between 2001 and 2005, only 181 received a Scholastic Audit during this time period. The current study focuses on these elementary schools ($N = 181$) as the unit of analysis.

Seventeen repeat incidences of administered Scholastic Audits were removed from the current study, for an investigable sample of 164 elementary schools ($n = 164$). While the SISI contains nine standards and 88 indicators grouped into three domains of school improvement, the current study focuses on only the first three standards of the SISI, commonly referred to as the Academic Performance standards: Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction.

Descriptive statistics, psychometric analysis, and multiple regression were employed in the current study to analyze the data. In terms of multiple regression, both simultaneous multiple regression and hierarchical multiple regression were used. Due to the nature of the secondary database, no missing or incomplete data sets were present. Three types of variables were incorporated into the current study: demographic control variables, independent variables, and dependent variables. Regarding demographic control variables, the researcher used the following: Year of Audit, County vs. Independent School status, Appalachian status, Size of School (4th grade), Percent White (4th grade), Percent Title I (4th grade), Percent Gifted (4th grade), Percentage Free/Reduced Lunch (4th grade), Percent Non-Disabled (4th grade), and Voluntary versus Involuntary status. As previously mentioned, several demographic control variables were removed from the regression model after being deemed not significant through the correlation matrix results (Dr. Douglas Smith, personal communication, April 29, 2016). The Academic Performance standards, the first three standards of the SISI—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—constituted the independent variables for the study. Finally, the elementary science component of the

Academic Index represented the demographic variable, measured during the fourth grade year of elementary school. While the Academic Index comprised 90.25% of the overall Accountability Index, elementary science accounted for 14.25% of the Academic Index.

Upon the completion of descriptive statistics for the study, factor analysis was conducted on the Academic Performance standards: Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction. For each of the three Academic Performance standards, factor analysis yielded a single factor for loading the respective indicators, accounting for 57.43%, 56.11%, and 52.35%, respectively. Cronbach's (1951) alpha was used in the current research to determine the internal reliability of the factor analyses. Cronbach's coefficient alphas of .874 for Standard 1, Curriculum, .879 for Standard 2, Classroom Evaluation/Assessment, and .865 for Standard 3, Instruction, confirmed a high degree of internal reliability for each single factor structure. In terms of validity, inter-scale correlations were assessed among the three Academic Performance standards, with r ranging from .681 to .811, and thus confirming the external criterion validity. Much like the previous research in this set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010), the psychometric analysis of the secondary database of Scholastic Audit data further confirms the efficacy of the SISI standards, as proposed by the KDE.

To assess the relationships presented in Figure 1, both simultaneous multiple regression and hierarchical multiple regression were employed in the current study. For Research Question 1, which examined the impact of demographic factors on the three Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—and the elementary science component of the Academic Index, the demographic factors had a significant effect on all

four simultaneous multiple regressions. In fact, demographic factors explained 62.6% of the total variance on the elementary portion of the Academic Index. In contrast, the effect of demographic factors on the three Academic Performance standards was considerably smaller, explaining 13.1%, 16.7%, and 17.3%, respectively.

Regarding Research Question 2, which sought to explain the relationship of the three Academic Performance standards on the Academic Index, Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction, had a combined impact to account for 28% of the total variance in the elementary science component of the Academic Index. However, only Classroom Evaluation/Assessment yielded a significant influence on the Academic Index, $\beta = .349, p \leq .01$.

For the final empirical research question, the researcher employed a hierarchical multiple regression model to explore the impact of the three Academic Performance standards on the elementary science portion of the Academic Index, while controlling for demographic factors. In the final model of the hierarchical regression equation, Standard 2, Classroom Evaluation/Assessment, in addition to five demographic factors (Year of Audit, Percent White, Percent Gifted, Percentage Free/Reduced Lunch, Voluntary versus Involuntary status), had a combined effect size of .680 on the Academic Index—a strong impact for elementary school educational research.

Aligning with the purpose of the current study, the central research question proffers the following: To what extent is Academic Performance (Standards 1-3 from Kentucky's *Standards and Indicators for School Improvement*) related to elementary school science accountability scores? The final hierarchical multiple regression model confirmed a moderate impact for alterable classroom practices on the Academic Index, when controlling for demographic factors. Yet, ultimately, in accordance with the

organizational systems theory of Senge (1990), the findings from the current study provide an optimistic, overarching outlook for educators that alterable classroom practices, and other mutable constructs, can influence enhanced academic outcomes for Kentucky students.

CHAPTER V

DISCUSSION AND CONCLUSIONS

Introduction

In the early 1980s, *A Nation at Risk* (National Commission on Excellence in Education, 1983) proved a watershed moment for educational reform, hyper focusing the national spotlight on the dismal state of the American education system, and eliciting immediate public cries for wholesale change. With this seminal document as the backdrop, and the battle cry for introspection at all levels of education, Kentucky began its own call to action with the Kentucky Education Reform Act of 1990 (KERA). KERA was a groundbreaking piece of legislation, the result of a lawsuit brought to the state by 66 Kentucky school districts that sought to remedy the inequity in educational funding, to level the playing field in governance, and to provide enhanced educational accountability outcomes for vested stakeholders. Steffy (1993) lauded KERA as “the most comprehensive, innovative reform legislation ever passed by any state in recent history” (p. xiii).

With a new accountability model in place through KERA, the Kentucky Department of Education (KDE) was tasked in determining the overall status of each school through the Kentucky Instructional Results Information System (KIRIS). KIRIS employed standardized tests to measure academic outcomes and represented the first statewide system in America to enact performance-based accountability (Stecher et al., 1997). Eight years later, with questions of validity and reliability linked to KIRIS, the

KDE moved forward with the Commonwealth Accountability Testing System (CATS) to assess overall school health (Stecher et al., 1997). CATS calculated school quality through an Accountability Index consisting of content area assessments, norm-referenced assessments, and additional measures (graduation rates, attendance, etc.). While CATS improved upon the reliability of KIRIS, the new accountability system still used several features of its predecessor, including the challenge for schools to achieve a score of 100 (*proficiency*) on the 140-point Accountability Index (Todd, 2010). To meet this lofty goal of accomplishing a score of 100 on the Accountability Index, the KDE (2000b) developed the *Standards and Indicators for School Improvement* (SISI) and the Scholastic Audit to guide schools toward proficiency.

Consisting of 88 indicators across nine standards, the SISI represented a standards-based guide for whole school reform and improvement. The SISI standards were divided among three distinct domains for school improvement: (a) Academic Performance, (b) Learning Environment, and (c) Efficiency. For the current study, the researcher focused on the three standards that constituted the Academic Performance domain: Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction. In as far as the current study, Academic Performance refers to the first three standards of the SISI, as outlined by the KDE, and not as a measure of school and district achievement. As cited from the *School Level Performance Descriptors and Glossary for Kentucky's Standards and Indicators for School Improvement* (KDE, 2008), KDE provided the following definitions for the first three standards of the SISI:

Standard 1: The school develops and implements a curriculum that is rigorous, intentional, and aligned to state and local standards. (p. 3)

Standard 2: The school utilizes multiple evaluation and assessment strategies to

continuously monitor and modify instruction to meet student needs and support proficient work. (p. 10)

Standard 3: The school's instructional program actively engages all students by using effective, varied, and research-based practices to improve student academic performance standards. (p. 18)

Based on scores from the Accountability Index, schools were divided into three categories of performance: (a) meeting their goal, (b) progressing, or (c) needing assistance. Schools that failed to meet proficiency, deemed as "needing assistance" by the KDE, were mandated to conduct a Scholastic Audit to help specifically address areas for growth. With 88 indicators across nine standards and subsequent three domains, the SISI document assisted struggling schools to glean insight and served as the guiding force behind the Scholastic Audit process and consequently whole school reform. In further measuring overall school quality, KDE used a subclassification system for schools designated as "needing assistance." Level 3 schools were the lowest one-third of the schools falling into the "needing assistance" classification and were forced to conduct a Scholastic Audit. Level 2 schools were the middle one-third of those schools "needing assistance" and required to use the Scholastic Review process. Finally, the highest one-third of schools "needing assistance" conducted an internal self-review. Scholastic Audit teams were composed of specific representatives trained by the KDE from each following vested educational party: a parent, a teacher, a school administrator, a district administrator, a university faculty member, and a Highly Skilled Educator (B. C. Ennis, 2007). While many schools were mandated to conduct a Scholastic Audit, based on their category of "needing assistance," successful schools could also volunteer to participate in a Scholastic Audit—a reflection of their desire for continuous improvement. The current

study includes cases from both voluntary and involuntary categories.

Within the three domains of whole school reform outlined in the SISI, the first three Academic Performance standards represented the reform measures that occurred within the classroom. In terms of the first SISI standard, Curriculum, Farenga et al. characterized this as the construct through which teachers and students interact (as cited in Bybee, 2002). DuFour et al. (2006) encouraged the use of Professional Learning Communities to facilitate enhanced curricular conversations among educators. The KDE proffered seven indicators for Curriculum to help guide schools in reform measures—either through proactive measures taken by a respective school or through guidance from Scholastic Audit team feedback.

Regarding the second SISI standard, Classroom Evaluation/Assessment and its eight indicators, the KDE (2004b) encouraged the implementation of a rigorous assessment design program that afforded opportunities for teachers to foster continuous student growth toward proficiency (p. 10). Assessment and evaluation have a profound impact not only on monitoring student learning, but, as a harbinger for future modifications to the curriculum and instruction as well. Stiggins (2008) noted assessment as a means for gathering information to drive future instruction (p. 1). Research indicates a summative assessment program does not meet all the diverse needs of learners. With its ability to provide frequent and informal diagnostic information to teachers regarding student learning, formative assessments supplemented summative evaluations in many classrooms.

In terms of the final SISI Academic Performance standard, Standard 3, Instruction, the eight indicators of Instruction outlined an alignment with the innovative educational thought processes of the period (KDE, 2008). Parallel with the spirit of many of the SISI indicators for Instruction, Zemelman et al. (2012) postulated a series of seven “best

practices” for instruction, beginning in the early 1990s. With best practices including a “gradual release of responsibility” and “collaborative activities,” educational methodology began the seismic paradigm shift from teaching to learning—from a teacher-centered classroom to a student-centered classroom. In fact, to reflect this migration of thought in the classroom from teaching to learning, Bray and McClaskey (2015) encouraged educators to call students learners—with a keen focus on the active process of self-direction as learners rather than the passive process of knowledge reception as students.

Another important measure for the assurance of quality instruction in the classroom, the KDE (2008) incorporated the recruitment and development of highly qualified teachers as an indicator for sound instructional practices. As the following educational axiom suggests, “a teaching degree will no more ensure a quality teacher as a driving license will ensure a quality driver.” Both Hattie (2009, 2012) and Zemelman et al. (2012) cited the paramount value of a quality teacher in the classroom for enhanced academic outcomes for students. Perhaps highlighting the recursive nature of educational theory, both the SISI document and current instructional best practices maintain that students driving their own learning, with the facilitation and guidance of highly qualified educational professionals, truly catalyzes student achievement and student success.

Through an investigation of the secondary Scholastic Audit database maintained by the KDE, the current research aligns with the set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) directed by Drs. Miller and Smith of Western Kentucky University. The current study seeks to examine the effects of the Academic Performance standards, alterable classroom practices, on the elementary science component of the Academic Index, a subset of the overall Accountability Index. The central research question for this study is the following: To what extent is Academic

Performance (Standards 1-3 from Kentucky's *Standards and Indicators for School Improvement*) related to elementary school science accountability scores?

In the remaining sections of this chapter, the researcher includes **The Study in Brief, Discussion and Analysis, Recommendations, and Conclusions.**

The Study in Brief

The current study is a quantitative analysis of Scholastic Audit data from a secondary database maintained by the Kentucky Department of Education (KDE) and demographic school factors from the Kentucky Performance Report (KPR). The purpose of this research is to examine the following: (a) possible relationships of demographic factors on the *Standards and Indicators for School Improvement* (SISI) Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—and the elementary science portion of the Kentucky Academic Index; (b) possible relationships between the Academic Performance standards and elementary science achievement in Kentucky; and (c) possible relationships between the Academic Performance standards and the elementary science component of the Academic Index, while controlling for demographic factors. As mentioned previously in this chapter, the central research question is the following: To what extent is Academic Performance related to elementary school science accountability scores?

The researcher examined secondary Scholastic Audit data collected by KDE-trained teams of educators and parents. Scholastic Audits were conducted at low-performing Kentucky schools to identify areas of growth for schools in the annual effort to progress toward a score of 100 on the Accountability Index by 2014. As the guiding document for the Scholastic Audit process, the SISI afforded schools nine standards with

88 indicators across three domains of school reform to assist in approximating exact areas of improvement. For each of the 88 indicators, the KDE employed a four-point behavioral scale to measure performance toward a specific indicator. While the measurement of this KDE-designed four-point scale is 1, 2, 3, 4, the current research converted these scores to a 0, 2, 5, 7 scale to reflect the more difficult process for schools to move from a 3 to a 4 on the scale (equivalent to a score of 100) on the Accountability Index (McKinney, 2007).

For the current study, the school was the unit of analysis. Specifically, the current research examines Kentucky elementary schools ($N = 181$) from 2001-2005, including both those mandated to conduct a Scholastic Audit and a voluntary group of high-performing schools that completed Scholastic Audits as a method for continuous improvement. Thus, both voluntary and involuntary schools that conducted a Scholastic Audit are included in the current research. Of note, since cases of repeat audits exist, the researcher eliminated these 17 cases of subsequent audits from the sample—with the assumption that schools who completed an additional audit in the following year would skew the validity and generalizability of the data. Thus, the investigable sample for the current study is 164 Kentucky elementary schools ($n = 164$) that completed one Scholastic Audit between 2001-2005. During the five rounds of Scholastic Audits between 2001 and 2005, the following cases were conducted each cycle: 56 audits in 2000-2001, 17 audits during 2001-2002, 76 audits in 2002-2003, 5 audits during 2003-2004, and 10 audits in 2004-2005 ($n = 164$).

For the current study, three types of variables were analyzed for possible relationships: demographic control variables, independent variables, and a dependent variable. In terms of the demographic control variables, multiple demographic factors were included in each dissertation from the current set (B. C. Ennis, 2007; McKinney,

2007; Saravia, 2008; Todd, 2010), as overseen by Drs. Miller and Smith, and these factors were found to yield a considerable and significant impact on the demographic variables in the research. While each of the previous dissertations used a subset of demographic factors from the overall list first presented by L. S. Ennis (2002), the current research employed a more comprehensive list of demographic controls for initial exploration of relationships among variables. These demographic control variables included the following: Size (SIZE), Ethnicity (%WHITE), Appalachian Status (APP), Percentage Free/Reduced Lunch (%FRED), Percentage Non-Disabled Students (%NDAB), Percentage LEP Students (%LEP), Percentage Migrant Students (%MIG), Percentage Gifted/Talented Students (%GFT), Independent/County School District (CTY), Year of Audit (YEAR), Percentage Title I Students (%TITLE1), Percentage Female Students (%FEM), and Percentage Extended School Service Students (%ESS). While the current set of demographic control factors were not an exhaustive list of the demographic variables that could affect school outcomes or measures of the three Academic Performance standards (from the Academic Index and the SISI, respectively), they aligned with the control variables used in the aforementioned set of dissertations. Additionally, only fourth grade students and subsequent demographic descriptors were included in the current study, since the elementary science component of the Academic Index was administered during the fourth grade year.

For the current research, the Academic Index, specifically the elementary science portion, constituted the dependent variable. Elementary science scores for the Academic Index were collected from the Kentucky Performance Report (KPR). The three Academic Performance standards from the *Standards and Indicators for School Improvement (SISI)* document—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and

Standard 3, Instruction—comprised the alterable independent variables (cf. Bloom, 1980). To obtain scores for the three Academic Performance standards, the researcher accessed the secondary database of Scholastic Audit data, maintained by the KDE.

In terms of analysis for the current study, the researcher used a correlational design. Descriptive statistics were reported for the independent variables, including the demographic control variables. Besides the current set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) and an analysis of an updated rural construct (a measure of place) by Smith and Miller (2013), virtually no studies have been conducted to assess the efficacy of the Scholastic Audit and the SISI standards—including by the KDE. Therefore, for the independent variables of the three SISI Academic Performance standards and their subsequent indicators, the researcher employed the psychometric analysis techniques of Factor Analysis and Cronbach's (1951) alpha to gauge the possibility of data reduction and to assess the internal reliability of these reduced factors, respectively. Regarding Factor Analysis, Costello and Osborne (2005) cautioned against using principal component analysis (PCA) as the default setting during factor analysis, as it is the case in many statistical software packages, since PCA often inflates factor scores. Therefore, the researcher utilized maximum likelihood (ML) exploratory factor analysis for the current study. Additionally, eigenvalues and scree plots were used in the current research to confirm any underlying factor structures revealed by the ML analysis. Costello and Osborne (2005) encouraged the use of a factor rotation method after extraction and retention methods had been confirmed. For the current study, the researcher used an oblique rotation method, based on the clarity and reproducibility cited by Costello and Osborne (p. 3). In terms of the internal reliability of subsequent data-reduced factors, Vogt (1999, p. 64) suggested that Cronbach's (1951) alpha values of 0.7 or higher as statistically

significant.

In the current study, multiple regression was employed as the main form of analysis to assess relationships between one or more independent variables on a dependent variable, while controlling for demographic factors. Vogt (1999) noted that regression analysis attempts to answer the following question: “What values in the dependent variable can we expect given certain values of the independent variable(s)?” (p. 240). For the current research, the statistical software package SPSS was used to calculate the multiple regressions that sought to answer the three empirical research questions, regarding the impact of the three SISI Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—on the elementary science portion of the Academic Index, while controlling for demographic factors. In terms of specific multiple regression methodology, simultaneous multiple regressions were employed for Research Questions 1 and 2, while hierarchical multiple regression was used for Research Question 3.

Both the Institutional Review Boards at the University of Louisville and Western Kentucky University approved this study as Not Human Subjects Research (NHSR). All accompanying documentation of this decision is included in Appendix B.

Discussion

In this section of the current Chapter V, the researcher looks separately at how the results of the current study relate to prior work in the field for Descriptive Statistics, Psychometric Analysis, and Empirical Research Questions.

In examining the history of accountability for both Kentucky and the nation, and the impact of best practices in the classroom for curriculum, classroom evaluation/assessment, and instruction, a wealth of educational research exists that assisted

the researcher in composing a review of literature to guide the current study. Levin (1974) outlined the varying components and systems behind accountability. As mentioned in Chapter II of this work, Levin (1974) proposed the following four distinct systems of accountability: as performance reporting, as a technical process, as a political process, and as an institutional process (pp. 363-364). With performance reporting serving as the traditional understanding of accountability in education, Levin found this construct of performance reporting was easily understood—with its focus on outcomes—by educators.

As for alterable classroom practices, Zemelman et al. (2012) provided an abundance of best practices that could be categorized under the three *Standards and Indicators for School Improvement (SISI)* Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction. Zemelman et al. composed their work on best practices in the early 1990s, which paralleled in decade the developmental timeline of the Kentucky Education Reform Act of 1990 and the SISI document. Thus, several instances of best practices research are discussed through their incorporation into the indicators for the respective SISI standards. While Zemelman et al. (2012) discussed best practices in these three aforementioned areas, DuFour et al. (2006) provided insight into the collaborative, constructive process of Professional Learning Communities (PLCs)—teams of like-minded teachers—and the subsequent impact of PLCs on the classroom.

Through the SISI, the Kentucky Department of Education (KDE) delivered to Kentucky educators a complete model of whole-school reform that served as the guiding document for the Scholastic Audit process. Moreover, through the three Academic Performance standards of the SISI, Kentucky educators received a comprehensive framework of standards and indicators to help guide alterable classroom practices in

curriculum, classroom evaluation/assessment, and instruction.

Looking at the current field of educational research as it pertains to the Scholastic Audit and the SISI, several studies exist in the current set of dissertations that examine the Academic Performance standards and the Scholastic Audit process at the elementary level. McKinney (2007) examined Standard 7 (Leadership) at the elementary level, with Standards 1 (Curriculum) and 3 (Instruction) as mediating factors. Also, at the elementary level, B. C. Ennis (2007) examined SISI Standard 7 (Leadership) as mediated by Standard 6 (Professional Growth, Development, and Evaluation). Replicating the study of McKinney (2007), Todd (2010) also investigated Standard 7 (Leadership), with Standards 1 (Curriculum) and 3 (Instruction) as mediating variables, only this time at the high school level. These three dissertations all utilized Murphy's (2004) mediated instruction leadership model as a framework. In addition, Saravia (2008) studied the effects of Standard 4 (School Culture) and Standard 5 (School, Family and Community Support) for the elementary level. The Smith and Miller (2013) study looked at all nine of the standards within their respective groupings of three standards each in the broader areas of Academic Performance, Learning Environment, and Efficiency, but with a focus on the new rural (place) measure. Therefore, the extant field of educational research on the Scholastic Audits and the SISI has several exploratory studies at the elementary level and one at the secondary level. The current research is the only study to include all three Academic Performance standards to evaluate their individual and collective direct impact on the Academic Index, and the only research that examines only one content area (science) apart from the totality of the Academic Index.

Descriptive Statistics

Summary. Descriptive statistics were reported for the demographic control

variables (Table 1) and the dependent variable, the elementary science component of the Academic Index. The demographic control factors included the following: Size (SIZE), Ethnicity (%WHITE), Appalachian Status (APP), Percentage Free/Reduced Lunch (%FRED), Percentage Non-Disabled Students (%NDAB), Percentage LEP Students (%LEP), Percentage Migrant Students (%MIG), Percentage Gifted/Talented Students (%GFT), Independent/County School District (CTY), Year of Audit (YEAR), Percentage Title I Students (%TITLE1), Percentage Female Students (%FEM), and Percentage Extended School Service Students (%ESS). Additionally, descriptive statistics are included in the section below for the three Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction.

Analysis. For the schools who conducted a Scholastic Audit ($n = 164$), the average size of the fourth grade level was 60 students, with a minimum of 6 students and a maximum of 145 students. Since the elementary science component of the Academic Index was administered during fourth grade, only fourth grade students were included in school size. Regarding Percentage Female, the percentage of females in the current sample was 47.59%, and had no significant correlations with any other demographic factor. For ethnicity, 84.69% of the sample were white, comparable to the statewide average of 84.8% during the 2005-2006 school year (KDE, as cited in Saravia, 2008). For Title I school designation, 84.04% of the sampled schools were identified as Title I schools, well above the 71% statewide average for the 2005-2006 school year (KDE, as cited in Saravia, 2008). For the audited schools, participation in the Free and Reduced Lunch program was 61.20%, compared with the lower state average of roughly 50% (Miller et al., 2006). Due to the nature of Scholastic Audits, mandated for lower-performing schools in Kentucky, the higher percentage of Free and Reduced Lunch program participation aligned with state

and national trends.

Fram, Miller-Cribbs, and Van Horn (2007) noted the common conception of education as the great social equalizer. However, Fram et al. discovered through the research that this educational ideal was not necessarily the reality in schools. Braswell et al. characterized the academic achievement gap between poor and nonpoor children as “troublingly high” (as cited in Fram et al., 2007, p. 310). Fram et al. cited a litany of educational research studies noting that having higher rates of ethnic minorities and poverty in a school is correlated with diminished student achievement (p. 310). Fram et al. continued by stating that poverty dynamics might also play a role in lower student achievement among peer groups:

From a social justice perspective, we note that even “cultural” processes may ultimately reflect structural causes to the degree that an individual’s choices, beliefs, values, and behaviors are shaped by unequal access to resources and opportunities, institutional oppression, and processes of marginalization. One such structural cause is the quality of education itself. (p. 310)

Therefore, noting the combined impact of a variety of socioeconomic and social justice factors that contribute to lower student achievement and how these demographic forces play a role in the journey toward proficiency (a score of 100 on the Accountability Index) by 2014 is important. Of note, “failing” schools are clearly disproportionately those that are disadvantaged economically—further confirmation that progress on the equity dimension is a much more difficult process than progress on the excellence dimension. Therefore, this study also provides more confirmation for the thesis in Miller and Moore (2006) that the equity dimension (effects of economic disadvantage) is much more resistant to statewide (macro-level) progress than the excellence dimension (overall higher

achievement) where progress due to KERA is widely evident.

For the Appalachian demographic factor, less than half (.40) of schools received the distinction of Appalachian status. Appalachian region students account for nearly one-third of all students in Kentucky, with this region exhibiting characteristics of rurality, born from a focus on coal mining (Caudill, 1963; Smith, 2005, 2006). Appalachian status has significant correlations to Percentage Free and Reduced Lunch and Percentage Title I, both at the $p \leq .01$ level. For future research, the high correlations among the three aforementioned demographic control variables suggests further consideration for measures of sociodemographic factors since they are largely addressing the same construct. Regarding County versus Independent school districts, 91% of the schools audited were County schools, in comparison to the 9% Independent school district counterparts.

Students designated with Limited English Proficiency (LEP) was .66%, while Migrant students accounted for 2.29% of the sample ($n = 164$). While both descriptive statistics were insignificant figures for the sample as a whole, the percentage of LEP and Migrant students aligned with the 2005-2006 statewide average of 1.57% for both demographic factors (KDE, as cited in Saravia, 2008).

Of the students participating in the Extended School Services (ESS) program, 36.04% of the sample was designated as ESS, significantly higher than the statewide average of 21.6%. According to Saravia (2008), the ESS program is another indication of low-income students. The ESS program provides additional assistance for struggling students through before-school and after-school tutoring, typically with certified classroom teachers offering the services and then being reimbursed for their time through ESS funding. ESS yielded significant correlations with Percentage Free and Reduced Lunch and Percentage Title I, with a significant negative correlation to school Size, all at the $p \leq$

.01 level. Therefore, of note, smaller schools in the current sample had larger populations of students receiving ESS services, which negatively correlates with diminished Academic Index scores.

For Gifted students, the average Gifted identification for the sample ($n = 164$) was 13.78%, below the state average of 16.6% for the 2005-2006 school year (KDE, as cited in Saravia, 2008). Therefore, decreased percentages of students participating in the gifted program could have contributed to decreased elementary science scores on the Academic Index. However, the label of Gifted and its subsequent impact on student achievement in Kentucky could be misleading since students could be identified as Gifted for leadership (based on teacher recommendations) and not solely on scholastic aptitude (Dr. Vicki Riley, personal communication, May 7, 2016). Regarding the correlation of Gifted status to the Academic Index, the percentage of Gifted students was a moderate predictor of higher student achievement ($r = .506, p \leq .01$). For percentage of students who were characterized as Non-Disabled, the descriptive statistic mean was 85.73%. Thus, with 14.27% of students who were disabled, this figure exceeded the statewide average of 11%, as reported by the KDE in 2006 (as cited in Todd, 2010). The Percentage of Non-Disabled students had a significant negative correlation with Percentage Free and Reduced Lunch, $r = -.213, p \leq .01$; therefore, schools in the current sample who reported more Non-Disabled students were consequently likely to have a lower Free and Reduced Lunch Percentage (lower poverty rate).

Finally, for the Year of the Audit, most of the schools were audited early in the five cycles of the audits between 2001 and 2005, with an average audit year of 2001.37. Therefore, the significance of Year of Audit indicated that KERA was working and that achievement (across the state, and as reflected in schools audited) were doing better over

time so that a 2005 audited school was part of a higher statewide baseline and hence the positive correlation; later audits correlated with higher overall statewide achievement and the Year of Audit measure reflected this finding (Dr. Stephen Miller, personal communication, July 16, 2016).

Of note, 17 schools were removed from the current sample ($n = 164$) as repeat incidences of audits could skew the data. Also, in the current sample, 130 elementary schools volunteered to conduct a Scholastic Audit, while 34 schools were mandated by the KDE to conduct a Scholastic Audit. While other studies in the current set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) noted the large range among schools and the low mean on the combined measure of samples for the Academic Index scores (with some schools in the sample conducting mandated Audits as Level 3 Assistance Schools and some voluntary), this is the only study to date to provide descriptive statistics for both voluntary and non-voluntary schools. Of note, the mean score on the elementary science assessment for the schools forced to conduct a Scholastic Audit ($M = 52.89$) was lower than the mean of those elementary schools who volunteered for a Scholastic Audit ($M = 73.6$), reflecting the struggling nature of the lower-performing schools forced to conduct an audit.

Psychometric Analysis

Summary. For the current study, psychometric analyses included Factor Analysis and Cronbach's (1951) coefficient alpha for each of the three SISI Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction. Additionally, inter-scale correlations were conducted among the three Academic Performance standards. Per the development of the SISI document by the KDE, each of the nine SISI standards contains respective indicators that assess the same

respective construct. To date, no formal psychometric analyses have been performed on the set of SISI standards and indicators, apart from the current set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) chaired by Drs. Miller and Smith and an analysis by Smith and Miller (2013) focused on an improved measure of rurality (place).

Leading up to this study, in the current set of dissertations investigating the Scholastic Audit and the SISI document, only limited psychometric analyses were conducted on the Academic Performance standards. In the previous studies of both McKinney (2007) at the elementary school level and Todd (2010) at the high school level, the indicators for both Standard 1 (Curriculum) and Standard 3 (Instruction) loaded onto one factor, respectively. For the current study, the researcher employed eigenvalue calculations and scree plots, as well as the maximum likelihood analysis suggested by Costello and Osborne (2005), for the three sets of indicators in Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction), respectively.

For Standard 1, the seven indicators loaded on one factor, accounting for 57.43% of the total variance among the indicators. The one factor for Curriculum also yielded a high degree of internal reliability with a Cronbach's (1951) coefficient alpha of .874. The verification for Standard 1 aligned with the work of the KDE that the seven indicators for Curriculum should be assessed as one factor. For Standard 2 (Classroom Evaluation/Assessment), the eight indicators also loaded onto one factor, explaining 56.11% of the total variance among the indicators. No previous studies had confirmed the efficacy of the standard and indicators for Standard 2, Classroom Evaluation/Assessment. Additionally, the one factor solution for Classroom Evaluation/Assessment produced a high degree of internal reliability, with a Cronbach's coefficient alpha of .879. The

validation of Standard 2 confirmed the work of KDE that the eight indicators for Classroom Evaluation/Assessment should be measured as one construct. For the last Academic Performance standard, Standard 3 (Instruction) possessed eight indicators that loaded on one factor, accounting for 52.35% of the total variance. Instruction yielded a high degree of internal reliability, with a Cronbach's alpha of .865. Again, the verification of Standard 3 confirmed the work of the KDE that the eight indicators for Instruction should be considered as a unified construct.

Analysis. Adding to the work for the set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) and parallel investigation by Smith and Miller (2013), the current research confirmed the validity and high internal reliability for three of the nine standards of the SISI document: Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction. Specifically, this research confirmed the previous work of both McKinney (2007) and Todd (2010)—in elementary schools and in high schools, respectively—through the validation of Curriculum and Instruction in the current study. Of note, while other studies have examined specific Academic Performance standards from the SISI, such as Standard 1 (Curriculum) and Standard 3 (Instruction) by McKinney (2007) and Todd (2010), this is the *only* study to examine all three Academic Performance standards from the SISI document. Additionally, the current research is the only study to examine all three Academic Performance standards through the lens of the elementary science component of the Academic Index.

Since the Scholastic Audits, and the subsequent guiding document of the SISI, drive Kentucky whole school reform, validating and verifying the SISI standards and subsequent indicators through psychometric analyses was imperative. The current study provides validation and verification through ascertaining that each of the three Academic

Performance standards (Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction) possess sound—but not perfect—psychometric properties. For example, as noted in Chapter IV of the current research, Standards 2 and 3 yield less than ideal independence. This finding prefigures the later findings in the new SISR that Standards 2 and 3 represent separate constructs but are not as distinct as they should be for perfect psychometric properties (Dr. Stephen Miller, personal communication, July 14, 2016). With the daunting challenge of schools to reach proficiency (100) on the Accountability Index by 2014, the confirmation of the three Academic Performance standards and their subsequent indicators bolsters the Scholastic Audit process and affords educators a powerful vehicle for whole school reform through the *Standards and Indicators for School Improvement*.

Research Question 1

To what degree do demographic factors relate to Academic Performance (Standard 1, Standard 2, Standard 3) and to KCCT Science?

Summary. Four separate simultaneous multiple regressions were conducted in Research Question 1, to ascertain the impact of demographic factors on the three Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—and on the elementary science component of the Academic Index. The following demographic variables were included in the regressions: Year of Audit, County vs. Independent School, Appalachian status, Size of School (4th grade), Percent White (4th grade), Percent Title I (4th grade), Percent Gifted (4th grade), Percent Free and Reduced Lunch (4th grade), Percent Non-Disabled (4th grade), and Voluntary versus Involuntary status. Percent Female (4th grade), Percent Migrant (4th grade), Percent LEP (4th grade), and Percent ESS (4th grade) were removed

from the regressions after yielding not significant results through the correlation matrix (Dr. Douglas Smith, personal communication, April 29, 2016).

In terms of the impact of the demographic factors on the three respective Academic Performance standards, demographic variables accounted for very little influence on Standard 1 (Curriculum), accounting for just 13% of the variance, Adjusted $R^2 = .131$. Only Year of Audit ($\beta = .232, p \leq .01$) and Voluntary versus Involuntary status ($\beta = -.206, p \leq .05$) yielded significant effects. Likewise, the demographic variables accounted for very little variance on Standard 2 (Classroom Evaluation/Assessment), Adjusted $R^2 = .167$. Year of Audit was the only demographic factor that significantly affected Classroom Evaluation/Assessment, $\beta = .213, p \leq .01$. Finally, in terms of Instruction, demographic variables again accounted for a very small portion of the variance of Standard 3, with an Adjusted R^2 of .173. Only Voluntary versus Involuntary status yielded a significant impact on Instruction, with $\beta = -.181, p \leq .05$.

In examining the impact of the demographic factors on the elementary science portion of the Academic Index, demographic factors explained more than 62% of the total variance, with an Adjusted R^2 of .626. In the regression model, five demographic control variables yielded a significant impact on the Academic Index. Year of Audit had the greatest effect on the Academic Index, with $\beta = .314, p \leq .01$. Voluntary versus Involuntary status and Percentage Free and Reduced Lunch both had negative impacts on the Academic Index, with $\beta = -.286, p \leq .01$ and $\beta = -.257, p \leq .01$, respectively. Percentage White for fourth grade had a beta value of .187 at the $p \leq .01$ level, while Percent Gifted for fourth grade yielded a beta value of .175 at the $p \leq .01$ level.

Analysis. Regarding the results of the current study, the three Academic

Performance standards are each moderately affected by the demographic control variables included in the research. Demographic factors account for 13% of the variance on Standard 1 (Curriculum), with Voluntary and Involuntary status and Year of Audit each yielding significant effects. Additionally, demographic factors influenced Standard 2 (Classroom Evaluation/Assessment), accounting for nearly 17% of the overall variance. Year of Audit was the only demographic variable significantly affecting Classroom Evaluation/Assessment. For Standard 3 (Instruction), demographic factors accounted for 17% of the overall variance in Instruction, with only Voluntary versus Involuntary status yielding a significant effect.

The notion that any student because of demographics is receiving unequal treatment, regarding Curriculum, Classroom Evaluation/Assessment, and Instruction, is alarming. Capturing 30 years of research, Darling-Hammond contended that student outcomes are consistently affected by four main factors: (a) smaller school size (around 300-500 students), (b) smaller class size (particularly in elementary school settings), (c) challenging curriculum, and (d) more highly qualified teachers—each of which is unequally offered to minority groups, as opposed to their white counterparts (as cited in Roeder, 2000a, p. 5). Darling-Hammond proffered that unequal access among minority groups to educational resources, such as a quality curriculum and outstanding teachers, is more detrimental to student achievement outcomes than race (as cited in Roeder, 2000a, p. 5).

While Year of Audit was significant for Standard 1 (Curriculum) and Standard 2 (Classroom Evaluation/Assessment), indicating incremental gains in improvement under these two constructs in subsequent cycles of the Scholastic Audit administrations, research (Roeder, 2000a; Miller & Moore, 2006) indicates that these gains might be misleading

with respect to closing equity gaps. The overall statewide average is rising (KERA progress on excellence dimension), but progress on equity is painfully slow (Miller & Moore, 2006). Even with attempts to reorganize the accountability system to improve quality, as the KIRIS testing system was replaced by CATS due to issues of validity (Stecher et al., 1997), Roeder (2000a) found that nearly 90% of the variance in the 1999 scores on CATS could be predicted by the 1998 KIRIS scores. Thus, even with a modified accountability system, “high performers tend to remain high and low performers remain low” (Roeder, 2000a, p. 11). Equally alarming, Roeder (2000a) discovered that the performance achievement gap between advantaged and disadvantaged students was even more pronounced in urban school districts (such as Jefferson County and Fayette County) than in the rest of Kentucky—and more pronounced than racial achievement gaps (p. 8). Most disconcerted, Roeder (2000a) stated the following:

As a nation that prides itself on the importance of the individual and is characterized by the widely held belief that anyone can achieve, be successful, and contribute to society if they work hard and apply themselves, the findings are troublesome, especially as they relate to the one social institution that is supposed to provide equal opportunity for all—rich or poor, black or white; that is, our system of common or public schools. (p. 6)

Regarding the demographic impact on the Academic Index, the demographic factors accounted for more than 62% of the overall variance on the elementary science component of the Academic Index. Year of Audit, Percentage White, and Percentage Gifted all had a significant and positive impact on the Academic Index, while Voluntary versus Involuntary status and Percentage Free and Reduced Lunch both had a significant, negative impact on the elementary science portion of the Academic Index. Of note, the

significance of the Year of Audit supports (empirically) the rising statewide excellence, while all the other significant demographic variables are measures related to equity (Dr. Stephen Miller, personal communication, July 16, 2016).

In terms of the Year of Audit, the positive impact of this demographic factor on the Academic Index indicates successive statewide approximation toward the goal of proficiency (a score of 100) on the overall Accountability Index. In an era of educational reform, the finding of Year of Audit as significant is a hopeful measure that schools system wide can make progress on the Science Academic Index. This is a genuine success for KERA as policy efforts to improve school achievement for the state. This is in spite of socioeconomic factors that appear to preclude growth. Taking into account what schools can control and cannot control, schools are attempting to focus on the former, with apparent gains in elementary science. Roeder (2000a) cited a large increase in Academic Index scores between 1998 and 1999—the transition year between KIRIS and CATS—with subsequent years continuing this trend. While average school ranking did not fluctuate as much, schools statewide were still making progress toward the goal of proficiency. Of note, one limitation regarding the Year of Audit is the small *N* of sample cases in later years of the current sample (2004 and 2005 comprised 5 and 10 cases, respectively). However, the current research on Year of Audit supports the findings of other researchers (Miller & Moore, 2006; Roeder, 2000a, 2000b).

As seen in the results of the current research, Percentage Free and Reduced Lunch had a significant, negative impact on the Academic Index. While Percentage White had a significant, positive impact on the Academic Index, and while focused and sustained efforts of racial integration in Kentucky public schools continue to be worthwhile, Roeder (2000a) suggested that perhaps a greater investment by Kentucky in schools with larger

percentages of poverty would produce a greater return in student achievement (p. 13). Yet, expanded investments in schools cut against the current conservative mood of the country (lower taxes and reductions in “government” programs with the rationale of letting the individual make it on his/her own).

The problem in terms of policy is the temptation to look at such empirical evidence as separate findings for poverty and race. Still, this is not as simple as declaring a war on poverty and ignoring racial inequality. For example, Ogbu discussed how race and class were inexorably intertwined in the American context (as cited in Miller & Moore, 2006, p. 19). While race had a negligible impact on student achievement in their research, Miller and Moore (2006) proffered that this had less to do with effects on accountability and more to do with the profound progress made by Jefferson County Public Schools—the subject of their study—in racial desegregation. Thus, while the current study cites significant effects of both Percentage White and Percentage Free and Reduced Lunch on the Academic Index, more research is needed to reconcile the individual and combined influence of both demographic factors on student achievement in a variety of socioeconomic and racial contexts.

A related variable, Percentage Gifted, had a significant and positive impact on the elementary component of the Academic Index. While gifted students positively contribute to the elementary science portion of the Academic Index, Miller and Moore (2006) argued that the percentage of gifted students, and the subsequent positive influence of said students, indicated more than just individual student differences, as gifted students positively correlated with other socioeconomic and demographic factors. In the current research, Percentage Gifted negatively correlates with Percentage Free and Reduced Lunch ($\beta = -.390, p \leq .01$) and positively correlates with Percentage White ($\beta = .163, p \leq .05$).

While the current research focuses on students at the elementary level, the assertion of Miller and Moore still rings true: “thus the prominence of Gifted...in the middle and high school regressions is tempered by the link between social position and student aptitude that is part of the fabric of inequality in America” (p. 18).

As to the significant and negative impact of Voluntary versus Involuntary status on Standard 1 (Curriculum), Standard 3 (Instruction), and the Academic Index, it makes sense that low-performing schools—mandated by the KDE to conduct a Scholastic Audit—would not score as well on the aforementioned components of the SISI and consequently the Academic Index as their higher-performing, voluntary counterparts. Nevertheless, the inclusion of high-performing schools in the Scholastic Audit database led to the focus by KDE on the set of indicators—originally known as leverage points and later as variance points—that successful schools do differently.

However, that focus by KDE on individual indicators (from the 88 overall in the SISI) goes *against* the intent of the original development of the *Standards and Indicators for School Improvement* as the nine standards for whole school reform were focused on school improvement standard by standard. In that regard, the research agenda encapsulated by the set of dissertations and related analyses that demonstrate the validity and efficacy of the standards themselves for school improvement (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010; Smith & Miller, 2013; Riley, 2016) are aligned with the original intent of the SISI and their national reputation as among if not the most effective extant school improvement tool (see Mintrop, 2003; Mintrop & Trujillo, 2005).

Research Question 2

To what degree does Academic Performance (Standard 1, Standard 2, Standard 3)

relate to KCCT Science?

Summary. One simultaneous multiple regression model was utilized to answer Research Question 2. The Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—yielded a moderate effect on the elementary science component of the Academic Index, accounting for 28% of the overall variance in elementary science scores. In terms of individual impact, Standard 2 (Classroom Evaluation/Assessment) was the only Academic Performance standard from the SISI that had a significant impact on the Academic Index, $\beta = .349, p \leq .01$. Also, Standard 3 (Instruction) was close to yielding a significant effect on the Academic Index, $p = .073$. Contrastingly, Standard 1 (Curriculum) did not influence the Academic Index significantly, $p = .979$. Essentially, both Classroom Evaluation/Assessment and Instruction played a far greater role in accounting for the variance in the Academic Index than Curriculum.

Analysis. The current research extends the work of the prior set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) in examining the impact of the SISI document on student achievement, as measured through the Academic Index. Additionally, the current study specifically extends the work of McKinney (2007), at the elementary school level, and Todd (2010), at the high school level, regarding the paramount importance of the SISI Academic Performance standards on school improvement. Both McKinney (2007) and Todd (2010) found Standard 1 (Curriculum) did not yield a significant impact on the Academic Index. The results of the current research also aligned with these findings, as Curriculum was not close to being a significant predictor of Science Academic Index scores. While McKinney (2007) and Todd (2010) examined the impact of the Academic Performance standards on the entire

Academic Index at their respective school levels, the current study only examined the elementary science component of the Academic Index. Furthermore, both McKinney (2007) and Todd (2010) performed a mediated effects model of instructional leadership (Standard 7 mediated by Standards 1 and 3), but did not examine Standard 2. In contrast, this study looked at all three SISI standards from the Academic Performance block but did not investigate them as mediating Standard 7, Leadership.

Curriculum is the first standard of the SISI, with the KDE encouraging educators to begin with the standards and content being taught. However, Curriculum does not yield a significant impact on the elementary science portion of the Academic Index. Thus, regarding specific indicators, neither Roeder's (2000b) statement about the beliefs of educational reformers that "certain educational strategies, teaching practices and methods, and curricular innovations will greatly improve student achievement and school performance" (p. 3), nor "evidence that the curriculum is aligned with the *Academic Expectations, Core Content for Assessment, Transformations* and the *Program of Studies*" (KDE, 2008, p. 3, Indicator 1.1a) are supported in this study. In the current research, "curricular innovations" failed to influence significantly the elementary science component of the Academic Index. Perhaps the elementary schools in the current sample were not innovative enough in their curricular modifications to affect achievement in elementary science significantly. Or, perhaps, in terms of elementary science, it would appear it matters less *what* you teach as *how* you teach it (Standard 3), and even more importantly, *how* it is being assessed (Standard 2). Additionally, the finding of Curriculum as an insignificant predictor of Academic Index scores could be attributed to the type of science test being administered by the KDE. Without actual performance events or tasks for students to demonstrate mastery of science content, the KCCT science test therefore

becomes more of an indication of reading—and possibly mathematics—mastery than actual science content, with students having to *read* the science questions before applying any actual science content knowledge.

To address the current supposition regarding the lack of curricular innovation inherent in Standard 1, the examination of the mean, standard deviation, and range for the item (indicator) means of each of the three SISI Academic Performance standards is necessary. For Standard 1, Curriculum, the mean of the items is 2.27, while the item means for Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction, are 1.96 and 2.15, respectively. While Curriculum yields the highest mean, the range for its item means is .939, a restriction of range when compared with the range for item means of Instruction, 1.54, and especially in contrast to the range for item means of Classroom Evaluation/Assessment, 2.30. To further the aforementioned supposition, no item (indicator) for Curriculum possesses a mean above 3.00, while both Classroom Evaluation/Assessment and Instruction yield one and two indicators, respectively, that are above 3.00. In terms of standard deviation, Curriculum yields the highest deviation with 1.69, while Classroom Evaluation/Assessment and Instruction possess standard deviations of 1.51 and 1.55, respectively.

With respect to the mean of Curriculum, and taking into consideration the restriction of range among the item means and the high standard deviation, a “ceiling effect” exists for Curriculum, possibly caused by the standardization of curriculum (and subsequent curricular resources) mandated by the Kentucky Department of Education (KDE), according to the Kentucky Education Reform Act. As such, even poor-performing school districts are afforded the opportunity to meet adequately the curricular requirements of the KDE. Of note, and as observed in the above item means, standard deviations, and

item mean ranges, even if an indicator or standard represents important content, it will not correlate with enhanced academic outcomes if schools score virtually the same on that aforementioned indicator or standard.

In terms of Standard 2, Classroom Evaluation/Assessment yielded the greatest significant effect of the three Academic Performance standards on the Academic Index. The current research is the only study in the set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) chaired by Drs. Miller and Smith to examine the impact of Classroom Evaluation/Assessment on student achievement. Examining the integral role of assessment practices on achievement, Stiggins (2002) reported that educational reform movements have evolved over the past five decades to incorporate the three following components: “the articulation of higher achievement *standards*, the transformation of those expectations into rigorous *assessments*, and the expectation of *accountability* on the part of educators for student achievement, as reflected in test scores” (emphasis added by author; p. 759). Stiggins decried the current system of one-and-done assessment measures such as standardized testing being employed by many schools and classrooms when employed as the sole measure of student growth and performance:

We are a nation obsessed with the belief that the path to school improvement is paved with better, more frequent, and more intense standardized testing. The problem is that such tests, ostensibly developed to “leave no student behind,” are in fact causing major segments of our student population to be left behind because the tests cause many to give up in hopelessness—just the opposite effect from that which politicians intended. (p. 759)

Stiggins continued by stating that overarching, summative testing systems could not give educators the immediate, formative feedback requisite for continuous student

growth (p. 759). Stiggins (2002) outlined his vision of “assessment for learning”—not to be confused with formative assessments—by stating that assessment for learning included the learner in a process where teachers “use the classroom assessment process and the continuous flow of information about student achievement that it provides to advance, not merely check on, student learning” (p. 761). In the *fully functioning and operational level of development and implementation* descriptor for Indicator 2.1a of Standard 2 (Classroom Evaluation/Assessment), the KDE (2008) described classroom assessments as “frequent and are consistently used to ensure continuous student progress” (p. 10).

The influence of Classroom Assessment/Evaluation on elementary science in Kentucky could be attributed to a classroom assessment system predicated on the continuous advancement of student progress, eschewing the one-and-done nature of summative assessments as a sole harbinger of student growth in favor of more “assessment for learning” practices. Additionally, science lends itself to the constructivism inherent in assessment for learning—as “regular self-assessment” is another component of assessment for learning—as students take an active role in the recursive process of forming hypotheses, testing assumptions, drawing conclusions, and then making new hypotheses.

Interestingly, unlike the studies of McKinney (2007) and Todd (2010), Standard 3 (Instruction) did not yield a significant impact on the Academic Index; however, Instruction was close to being significant ($p = .073$). One possible explanation for Instruction not being significant in the current research, as opposed to the studies of McKinney (2007) and Todd (2010), is the presence of multicollinearity between Standard 2 (Classroom Evaluation/Assessment) and Standard 3 (Instruction). Vogt (1999) defined multicollinearity as the high correlation between two independent variables whereby it becomes almost impossible to determine their individual effects on the dependent variable

(p. 180). In the present study, Classroom Evaluation/Assessment and Instruction correlated very highly with $r = .811$ at the $p \leq .01$ level—which is higher than ideal for the two constructs to be sufficiently distinct, although “there is no magical value of correlation to define multicollinearity exactly” (Nunnally & Bernstein, 1994, p. 191). In the research of McKinney (2007) and Todd (2010), Standard 2 (Classroom Evaluation/Assessment) was not examined, suggesting that the significance of Standard 3 (Instruction) captured the high intercorrelation between Standards 2 and 3 demonstrated in the current research. Furthermore, regarding the impact on the dependent variable, Standard 2 had the more robust correlation with the Science Academic Index of the two aforementioned standards.

Another possible cause for this deviation from the results of McKinney (2007) and Todd (2010) could be that the current study focused only on the elementary science component of the Academic Index, while the other two aforementioned studies examined the entire Academic Index as the school outcome measure. Therefore, the following question arises: Are there specific instructional strategies inherent in science pedagogy that fall outside the spectrum of indicators for SISI Standard 3 (Instruction)? Essentially, the indicators listed by the KDE (2008) in the SISI document derived from more general best practices in pedagogy and are not content specific. Best practices in science such as Inquiry-Based Learning are not referenced specifically in any of the eight indicators for Instruction. Therefore, more research is warranted regarding the best practices of science and how these practices specifically affect student achievement—in terms of the Academic Index. Additionally, since most elementary school teachers are content generalists, perhaps the quality of science-specific instruction is not on par with secondary science counterparts. Again, more research is needed in the field of elementary science instruction and its subsequent effects on student performance.

Research Question 3

To what degree does Academic Performance (Standard 1, Standard 2, Standard 3) relate to KCCT Science, controlling for Demographic Factors?

Summary. Hierarchical multiple regression was utilized to investigate Research Question 3 to determine the effects of the three Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—on the Academic Index, while controlling for demographic factors. All significant demographic control factors—from the correlation matrix and Research Question 1—were included in the regression. Additionally, County versus Independent School status was included since it approached significance. Those excluded were Appalachian status, Size of School, Percent Title I, Percent Non-Disabled, Percent Female, Percent Migrant, Percent LEP, and Percent ESS—which were removed for yielding no significant results in the correlation matrix, nor on Research Question 1 (Dr. Douglas Smith, personal communication, April 29, 2016; Dr. Stephen Miller, personal communication, July 12, 2016). For all hierarchical regression models, the full and reduced model *F*-test computations were conducted using the sum of squares method. The researcher used computations to ascertain whether the hierarchical regression model improved during each subsequent step. During Step 1 of the hierarchical regression model, only demographic factors were added to determine the impact of these variables on the elementary science component of the Academic Index. During Step 2 of the hierarchical regression equation, the three Academic Performance standards were added to gauge whether the model improved in efficacy.

During Step 1, the demographic factors yielded a strong influence on the elementary science component of the Academic Index, accounting for 63% of the overall

variance with an Adjusted R^2 of .634. The following demographic factors were all significant at the $p < .01$ level: Year of Audit ($\beta = .310$), Percent White ($\beta = .200$), Percent Gifted ($\beta = .174$), Percent Free and Reduced Lunch ($\beta = -.266$), and Voluntary versus Involuntary status ($\beta = -.289$). County versus Independent School status was also significant, with a $\beta = .099$ at the $p < .05$ level.

After controlling for demographic variables, the three Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—were added to the hierarchical regression model to determine their collective impact on the Academic Index. The addition of the Academic Performance standards resulted in an increase in the Adjusted R^2 value of .046—from .634 to .680. All five of the aforementioned demographic variables—Year of Audit ($\beta = .262$), Percent White ($\beta = .213$), Percent Gifted ($\beta = .162$), Percent Free and Reduced Lunch ($\beta = -.203$), and Voluntary versus Involuntary status ($\beta = -.251$)—remained statistically significant at the $p < .01$ level. While neither Curriculum nor Instruction were significant variables for the model, Classroom Evaluation/Assessment was statistically significant on the Academic Index, with $\beta = .214$ at the $p < .01$ level. Additionally, County versus Independent School status did not maintain its significance, with $p = .082$. In this final equation of the hierarchical regression model, 68% percent of the total variance was accounted for in terms of the impact on the elementary science component of the Academic Index.

Analysis. Regarding Research Question 3 and the influence of the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—on the elementary science component of the Academic Index, while controlling for demographic factors, five

demographic control variables remained statistically significant at the $p < .01$ level throughout the two steps of the hierarchical regression model: Year of Audit, Percent White, Percent Gifted, Percent Free and Reduced Lunch, and Voluntary versus Involuntary status. In terms of the Year of Audit, this demographic control variable yielded a beta of .262. Therefore, as the accountability cycles progressed, viewed through the lens of the Scholastic Audit, accountability scores increased as it pertained to elementary science. While Year of Audit was limited to only a few cases in 2004 and 2005, this finding aligns with the work of Miller and Moore (2006) who noted that achievement scores had increased substantially under the Kentucky Educational Reform Act (KERA). However, Roeder (2000b) noted that while the educational reform package of KERA, and the subsequent accomplishments by these initiatives, are causes for celebration, much remains to be done to realize student achievement equitably across all demographics (p. 1), also a major conclusion of Miller and Moore.

Regarding the Percent White demographic factor, the findings of the current research support the previous findings of McKinney (2007) at the elementary school level and Todd (2010) at the secondary school level. Percent White yielded a beta of .213 on the elementary science portion of the Academic Index. While race plays a role in the current research findings, Miller and Moore (2006) noted that race played a negligible role in the regressions for their study. The discrepancy between the current study and that of Miller and Moore could be attributed to the demographics of the respective population, as the current research focused on the entire Commonwealth, while the work of Miller and Moore investigated Jefferson County Public Schools—a school district with extensive desegregation efforts in place.

For Percent Gifted, this demographic factor demonstrated a beta of .162 on the

elementary science component of the Academic Index. While Percent Gifted may be at an individual student level, Miller and Moore (2006) cited the correlations to other sociodemographic factors (such as race and socioeconomic status) which could account for its significance in the regression model. Therefore, the correlations to sociodemographic factors for Percent Gifted—not to mention how gifted children impact achievement scores through intellectual aptitude—would significantly influence student performance positively. Additionally, since students are identified as gifted during elementary school in all districts throughout Kentucky, the higher prevalence of a gifted population at a respective elementary school could serve to increase the achievement gap, as students who are identified as gifted receive special services catered to their area of gifted intelligence.

Regarding Percentage Free and Reduced Lunch, this demographic factor exhibited a significant, negative effect on the elementary science portion of the Academic Index, with a beta of $-.203$. In addition, the beta values for Percent Gifted and Percent Free and Reduced Lunch corresponded with historical data on trending student achievement for these two demographic factors. Extending the work of McKinney (2007) and Todd (2010) at the elementary and secondary levels, respectively, the current research determined the impact of Percent Free and Reduced Lunch specifically on the elementary science component of the Academic Index, rather than across the entire Academic Index. Additionally, the current study aligned with the findings of Miller and Moore (2006) that socioeconomic status significantly affected student achievement. Miller and Moore noted that while the excellence dimension of the educational reforms of KERA had been successful, the equity dimension of KERA still had much to be desired: “What seems obvious now is that presumptions underlying systemic school reform that are inherent in high-stakes accountability models are *not* powerful enough, by themselves, to erase

achievement gaps associated with socio-demographic factors” (emphasis added by author; p. 26). Additionally, the findings of the current research on Percent Free and Reduced Lunch paralleled the conclusions of Roeder (2000b), who stated the following:

Since the adoption of KERA in 1990, disadvantage continues to be the most substantial and significant determinant of the performance of Kentucky schools. In addition, school performance is stabilizing, indicating that the large gap between the haves and the have-nots is not likely to decrease or be reversed without significant policy changes. (p. 4)

Therefore, noted in the results of the current study, and the works of previous researchers (McKinney, 2007; Miller & Moore, 2006; Roeder, 2000b; Todd, 2010), socioeconomic status continues not only to affect student achievement significantly in a general sense, but, more specifically, fails to reduce and perhaps even widens the achievement gap that KERA sought to remedy through its sweeping educational initiatives aimed at equity.

Finally, in terms of the last demographic factor to influence the elementary science portion of the Academic Index significantly, Voluntary versus Involuntary status remained significant at the $p < .01$ level throughout the hierarchical regression model, with a beta of $-.251$. Due to the nature of Scholastic Audits, mandated for all of the lowest-performing schools in Kentucky, the KDE decided to conduct Audits in a sample of successful schools who volunteered to be audited. Therefore, the findings of the current research aligned with the logical belief that lower-performing schools, mandated to conduct a Scholastic Audit, would in turn score lower on student achievement measures such as the elementary science portion of the Academic Index—and thus yield a significant, negative impact on the KCCT science measures.

After taking into account the unalterable nature of the demographic factors that

account for more than 63% of the overall variance in the elementary science component of the Academic Index, educators should still feel hopeful that the alterable classroom practices being employed throughout Kentucky can still make a difference in the lives of students. In fact, through the hierarchical regression model employed during Research Question 3, an additional 5 percent of the variance in the elementary science portion of the Academic Index can be attributed to the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction. While Curriculum and Instruction are not statistically significant in the final hierarchical regression model, Classroom Evaluation/Assessment remains statistically significant at the $p < .01$ level. More specifically, with a one unit increase in Classroom Evaluation/Assessment practices, the Academic Index increases by 3.318 units on the science component, based on the unstandardized b coefficient. Also, the alterable classroom practice of Classroom Evaluation/Assessment yielded a positive beta value that exceeded all significant, unalterable demographic factors except for Year of Audit and Voluntary versus Involuntary status—a very favorable finding for educators. Of note, unlike the work of McKinney (2007) and Todd (2010), Instruction was not statistically significant—perhaps indicating multicollinearity between Standards 2 (Classroom Evaluation/Assessment) and 3 (Instruction).

If educators can continue to focus on what they can control, and are mindful of what they cannot control, the alterable classroom practices of the Academic Performance standards—while not as influential as several unalterable demographic factors—can still make a difference. And, for many schools under the Kentucky Education Reform Act, positive gains are sometimes made in inches and not in yards.

Recommendations

With the vast number of sociodemographic factors and socioeconomic elements at stake in the systems of educational reform, it would be understandable for educators to become disheartened. As time progressed toward the 2014 deadline for Kentucky schools to reach proficiency, Lindsey (as cited in Todd, 2010, p. 188) noted that only 37% of schools would reach proficient status, with 50% in the progressing category and 13% in the needs assistance classification. As both Roeder (2000b) and Miller and Moore (2006) asserted in their research, this ideal of proficiency becomes even more difficult as structures to support schools in reaching proficiency are plagued with inequity. Yet, there is hope for Kentucky educators. Within the current research, educators are afforded significant constructs through the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—to effect growth and change in Kentucky elementary schools. In the following paragraphs, implications for policy and practice are discussed, with suggestions for future research also included.

Implications for Policy

With the recent passage of the Every Student Succeeds Act (ESSA) through a bipartisan congressional effort, education agencies and school districts across the nation are eagerly anticipating, or perhaps anxiously awaiting, the actualization of this legislation—and what it will look like at the state, district, school, and classroom level. Upon signing the bill into law, President Barack Obama stated, “with this bill, we reaffirm that fundamentally American ideal—that every child, regardless of race, income, background, the zip code where they live, deserves the chance to make of their lives what they will” (U.S. Department of Education, 2015c). While a bold proclamation indeed, the aforementioned words of President Obama sound less like political rhetoric and more like

the language of House Majority Leader Greg Stumbo before signing KERA into passage—that is to say, hopeful, inspiring, committed (Steffy, 1993, p. 2). In the following paragraphs, the researcher will explicate several policy implications gleaned from the current research that add to the field of study for educational reform and align with several overarching measures laid out by the ESSA.

Signed into law fifty years after the Elementary and Secondary Education Act (ESEA) of 1965 (U.S. Department of Education), the ESSA outlines several sweeping reforms that are set to transform education: (a) holding all students to high standards of academics; (b) preparing all students to be college- and career-ready; (c) providing more children access to exemplar preschool services; (d) setting in place accountability measures to ensure all students—and schools—achieve and grow at a high level; (e) de-emphasizing standardized testing practices while simultaneously keeping parents and students aware of student progress; and (f) promoting innovative practices and investing in best practices (U.S. Department of Education, 2015c). Much like its ESEA predecessor, the ESSA is positioned more as a social equalizer—a civil rights law—than merely an education act. Unlike the ESEA, where both excellence and equity in education were suspect, the ESSA arrives at a time in the national education landscape where high school graduation rates are extremely high, dropout rates are at all-time lows, and more students are attending college than during any time throughout history (U.S. Department of Education, 2015c). Nevertheless, America is still plagued with issues of equity.

Through KERA, Kentucky planned to solve both issues plaguing education in the Commonwealth: to enhance achievement at very high levels (*excellence*) and to reduce achievement gaps in areas of socioeconomic status, race, and other demographic factors (*equity*) (Miller & Moore, 2006, p. 21). According to Miller and Moore, Kentucky

addressed the *excellence* construct with tremendous progress in student achievement; however, much like the nation as a whole, *equity* problems still exist. Therefore, in the following paragraphs, the author will first discuss the policy implications in light of both excellence and equity.

First, Year of Audit significantly contributed to the elementary science component of the Academic Index in the current research, with a standardized beta of .259. Essentially, as the Year of Audit progressed, schools in the current research subsequently progressed toward proficiency on the Accountability Index. Thus, mandating a Scholastic Audit for all Kentucky schools, much like Todd (2010) indicated, would provide equal access to whole school reform measures for every school—not just those schools who are struggling. As the backbone of the Scholastic Audit, the *Standards and Indicators for School Improvement (SISI)* provided nine standards and a respective 88 indicators of school improvement characteristics that all schools could use to approximate success. Several successful schools volunteered to conduct a Scholastic Audit, through which the KDE used to develop variance points—indicators of what successful schools do differently. If all schools conducted a Scholastic Audit on an annual or biannual basis, the KDE could maintain a complete and updated database of school reform characteristics—their level of implementation of the respective standards—that could be accessed not only by Kentucky schools, but by schools across the nation.

While the Scholastic Audits, and by default the *Standards and Indicators for School Improvement (SISI)*, were abandoned by the KDE due to the rising costs of implementation, such a policy implication—an annual or bi-annual Scholastic Audit for all schools—could be possible through the actualization of a recently developed school improvement document, the *School Improvement Scholastic Review (SISR)*, by a team of

researchers at Western Kentucky University. The SISR is a revision of the SISI and a replacement for the Scholastic Audit process. With the promising results of recent pilot testing of the SISR, former concerns about the Scholastic Audit and the SISI have been alleviated as the SISR is economically feasible—through its user-friendly implementation of teacher perceptual scales that can be completed in 30-45 minutes. Additionally, the SISR has comparable predictive power to the invasive and costly Scholastic Audit process (Dr. Stephen Miller, personal communication, July 18, 2016).

Second, Percentage Free and Reduced Lunch also significantly and negatively contributed to the elementary science portion of the Academic Index, again based on the standardized beta of $-.207$. For those schools mandated to conduct a Scholastic Audit, the Percentage Free/Reduced Lunch represented 73.47% of the sample population, while their voluntary counterparts had a Percentage Free/Reduced Lunch population of 57.99%. Ultimately, the Percentage Free/Reduced Lunch population played a significant role in the elementary science component of the Academic Index, with involuntary schools scoring a mean of 52.89, in comparison to the 73.36 scored by voluntary schools on the science portion of the Academic Index. Without fully disaggregating the data of the current research, the Academic Index mean of the 34 schools who were forced to conduct a Scholastic Audit would possibly be overlooked when considering the 69.11 Academic Index mean (elementary science) of the entire sample ($n = 164$) for both voluntary and involuntary schools. The findings of the current research aligned with the conclusions of Roeder (2000b) who noted, “underperforming schools were much more likely to have higher proportions of poor and minority children and to be located in school districts with lower levels of previous academic achievement and fewer financial resources prior to implementation of KERA” (p. 1). While investing heavily in poorer schools and school

districts has remained as one potential solution upon the passage of KERA, Miller and Moore (2006) suggested that “despite educators’ best efforts, extensive reforms and massive investments have *not* been able to diminish the effects of social position” (emphasis added by author, p. 21). Therefore, in terms of addressing the second component of the KERA solution—equity—toward school reform (Miller & Moore, 2006), one policy recommendation is the equitable assignment of students based on class—socioeconomic status—much like that of desegregation. Miller and Moore suggested the following: “neither housing patterns nor student assignment plans in this country would give credence to a groundswell of support for policies to reduce the concentration of the poor in the nation’s schools and classrooms. Yet...that strategy may be necessary if Americans are serious about closing achievement gaps” (p. 24).

Third, while “throwing money at a problem” is never an appropriate solution (Miller & Moore, 2006; Roeder, 2000b), perhaps an intentionality about equitable financial distribution can assist in overcoming the inequity of funding among the poorest school districts in Kentucky. Roeder (2000b) suggested the following, in terms of more equitable funding, to address issues of equity across Kentucky: (a) increased salaries for teachers, (b) increased teaching resources, and (c) enhanced physical facilities (p. 4). In education, there is the saying that “fair is not always equal.” While this adage applies to the classroom in making accommodation to students, it could also be applied to the funding provided to the poorest school districts in Kentucky. Roeder cited the poorest school districts as needing even more resources than their well-off school district counterparts: “To accomplish such a goal, more new dollars would have to be invested in the K-12 education system, and the SEEK program would have to be revised so that poorer districts acquire even more state resources” (p. 4).

Fourth, from the personal observations of the researcher over the past three years as a Cognitive Coach for the kid-FRIENDLY, Race To The Top-District Grant (through the Green River Regional Educational Cooperative), many poor school districts—often rural—have limited access to technology. Additionally, when rural school districts do have technological resources, the Internet bandwidth is often so poor that students are left without access to online resources. Of note, for the fifty states in the Union, Kentucky ranks last in average peak Internet connection speed (mbps) at 34 mbps (Fastmetrics, 2015). For comparison, Delaware ranks first in average peak Internet connection speed at 75 mbps, while states such as Mississippi and Arkansas connect at 37 mbps and 35 mbps, respectively. Essentially, students in Delaware can connect to the Internet at twice the rate of Kentucky. While this idiosyncrasy in Internet connection speed may appear as “much ado about nothing,” many teachers, upon finding that they cannot expeditiously connect to the Internet, will abandon using technology in the classroom in favor of other methods. Therefore, Kentucky must find a way to appropriate technology funds to the poorest school districts more equitably and allow all Kentucky children the ability to access these technological resources—or rob them of the opportunity to be fully prepared for college- and career-readiness in a globalized society.

Finally, since the decline of Professional Learning funding by Kentucky in the late 2000s, schools and school districts across the state have been hard pressed to give educators the training and development requisite for continuous improvement toward best practices. Since this development, many Kentucky school districts have elected to provide all professional learning opportunities in-house. For school districts like Jefferson County Public Schools, Fayette County Public Schools, and Daviess County Public Schools, to name a few of the larger school districts, maintaining an internal professional learning

program has not constituted a serious problem—as many of these school districts contain a wealth of highly qualified teachers and administrators who will share current pedagogical practices with peers. However, for the poorer school districts in Kentucky, many so small that the entire school district consists of one to two schools (K-8 or K-12 schools) and where the entire school district scores for a grade level or content area rest in the hands of one or two teachers, in-house professional learning opportunities are not as frequent or feasible. Roeder (2000b) noted that “if the key to successful learning, especially for poor children, is teaching, how can underperforming and disadvantaged schools improve and increase their teaching resources?” While Roeder suggested his own solution through a Kentucky Teaching Corps, this researcher believes that the answer could be much simpler. To improve the pedagogical prowess of current teachers, Kentucky must make a sustained and intentional effort to provide all school districts—especially the poorest districts—with the much needed funding for educators to attend professional learning opportunities in other districts and states—or, at the very least, the ability to hire consultants to train and coach educators in a respective school district.

Ultimately, through the dual lens of equity and excellence, the aforementioned proposals for policy could drastically improve Kentucky education, even among the poorest school districts. Roeder (2000b) proffered that “a significant, comprehensive, and long-term investment of educational resources in the poorest-performing school systems may help achieve higher levels of equitable excellence” (p. 5). Through a mandated Scholastic Audit for all schools, to ensure excellence, and a more balanced distribution of resources (such as technology, Internet bandwidth, and Professional Learning funds), to promote equity, Kentucky schools could begin to actualize fully the KERA belief that all children can learn at high levels.

Implications for Practice

In the current research, the impact of the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—on the elementary science component of the Academic Index was assessed, while controlling for demographic factors. In the current research, Standard 2 (Classroom Evaluation/Assessment) had a significant impact on the elementary science portion of the Academic Index, while Standard 3 (Instruction) was nearly significant and Standard 1 (Curriculum) was not significant. While much of the policy implications above addressed the impact of demographic factors on the Academic Index from a macro-level, the following paragraphs focus more on the influence of the Academic Performance standards—and respective indicators—on the Academic Index, explicitly for the classroom practitioner (such as teachers and principals).

First, since the current version of the Kentucky science assessment does not include performance tasks, Kentucky schools should truly focus on incorporating science across content areas. In the elementary grades, where nonfiction reading and math are often infused with high-interest science topics and graphs, respectively, the incorporation of science into subjects such as reading and math seems a natural fit for classroom teachers looking to enhance engagement and subsequent achievement with science content. From the personal experience of the author as an elementary science teacher, many students would return to school the following day pleading to know the results of a science experiment. This overwhelming curiosity from students simply did occur with rote memorization activities or worksheets. Additionally, with the recent implementation of the Common Core State Standards, where nonfiction reading is a primary focus, educators would benefit from the marriage of English-Language Arts (ELA) and science topics.

Furthermore, with the recent appointment of a new education commissioner, Dr. Stephen Pruitt (a founding father of the Next Generation Science Standards), reforms to the current Kentucky science assessment are impending, with performance tasks and other innovative science assessment practices—which would necessitate stand-alone science classes in the elementary grades—unquestionably in play for future iterations. While science should be emphasized during reading instruction, Romance and Vitale (1992) suggested the converse also had mutually beneficial effects. Romance and Vitale found that implementing applied reading skills during focused science instructional time enhanced student achievement, attitude, and self-confidence in both science and reading (p. 550). Therefore, in all instances, and specifically science, learning should not be taking place in silos for the elementary classroom teacher. Through enhanced engagement and student achievement, the incorporation of science instruction into other content areas (and other content areas into science) offers a wealth of learning opportunities for students.

Second, in terms of science instruction, Kentucky elementary schools should implement stand-alone science classes across grade levels—not just during the year of the assessment. One year of science instruction out of the six possible years for elementary students not only robs children of the experiments and activities that make learning fun, but, also, sends the subconscious message that science is somehow less important than other core subjects. Additionally, science instruction and learning should be engaging, not exclusively from a textbook. In terms of science instructional strategies, hands-on, inquiry-based science has been noted to increase student achievement (Aguiar et al., 2010; Lee et al., 2010). Aguiar et al. (2010) also called for professional development for science teachers that enhanced teacher opportunities to address science topics, tracking common science questions of students and developing best practices for responding to these

questions. Shymansky, Kyle, and Alport (1983) noted that children made significant strides in process skill development, attitude toward science, and achievement when exposed to innovative science programs (p. 401). In essence, science fosters critical thinking—process skills development (Shymansky et al., 1983). Science enhances engagement and achievement (Romance & Vitale, 1992). And, science precipitates a love of learning. Therefore, if Kentucky genuinely wants students who are college- and career-ready, whatever future path or track, the incorporation of science across content areas in the curriculum is paramount to future success in a 21st century, globalized economy.

Third, Standard 2 (Classroom Evaluation/Assessment) was found to have a significant, positive impact, accounting for nearly 35% of the variance on the elementary science component of the Academic Index. Therefore, more professional learning is needed for teachers in the best practices of assessment program implementation in the classroom. According to the National Research Council (2000), formative assessments open a world of possibilities regarding the interchange of information between student and teacher through a “learner-friendly” format that allows teachers to make critical and timely adjustments to instruction (pp. 24-25). Additionally, Baeder (2016) stressed the importance of formative assessment in the classroom, but lamented the number of educators who use formative assessment: “it’s not so much that formative assessment has been tried and found wanting; it has been found difficult and not tried” (Chesterson, as cited in Baeder, 2016, p. 1). Baeder noted the inefficiency in formative assessments, with teachers traveling home after the school day carrying 120 exit slips (a formative assessment technique) in their pocket. Furthermore, since administrators experience this inefficiency in collecting formative assessment data from teachers, the methodology of formative assessments is rarely either practiced or monitored. In response to the apparent

inefficiency of use of formative assessments by teachers and administrators, Baeder suggested leveraging technology in a powerful way to cure the inefficiency inherent in formative assessment practices. Ultimately, the use of formative assessments in the classroom is a practice tantamount to approximating mastery; therefore, through the planning and implementation—with the assistance of technology—of formative assessments, educators are afforded a powerful and timely tool for continuous student improvement.

Finally, Kentucky educators should engage in a professional learning community (PLC)—even in the smallest, poorest school districts. DuFour et al. (2006) established the PLC model as a vehicle for collaborative development around student learning. In terms of assessment, PLCs assist educators in developing a shared vision of proficiency for a topic, develop tasks to teach mastery of this standard, create assessments that measure this shared vision, and reflect intently on the level of student learning regarding the aforementioned standard. For larger school districts, the establishment of PLCs is facilitated either by many teachers—in one grade level or content area—in a specific school or many teachers in several schools across the district—intra-school meeting opportunities. For smaller school districts, PLCs can be established with other local school districts or through accessing learning networks created by educational cooperatives (such as the Ohio Valley Educational Cooperative or the Green River Regional Educational Cooperative). Ultimately, the implementation of PLCs in the school and school district let educators talk about teaching and, more important, to discuss the effect of teaching in student learning.

For Kentucky to rise to the challenge of providing a world-class education for all students, educators must realize the necessity of giving learners the requisite tools to be

fully college- and career-ready. Through the integration of science across content areas, besides stand-alone science classes, at the elementary level, Kentucky students will have the formative foundation needed to succeed in a 21st century globalized economy. Additionally, for teachers to rise to this challenge in practice, educators must be engaged in continuous professional conversations around what really matters: student learning. Professional Learning Communities serve as the vehicle to drive these conversations regarding learning. Only when educators truly make this arduous paradigm shift from teaching to learning—from “I taught it” to “They got it”—will students be able to assume the mantle of 21st century, globalized learners.

Implications for Future Research

Since the inception of the Kentucky Education Reform Act (KERA) of 1990, Kentucky has been at the forefront of the nation in school reform measures during the accountability era of American education. Ever moving forward, Kentucky produced the *Standards and Indicators for School Improvement (SISI)* as the guiding document for whole school reform and the subsequent Scholastic Audit process. As the passage of the Every Student Succeeds Act (ESSA) in 2015 coincided with the appointment of a new Kentucky Commissioner of Education, Dr. Stephen Pruitt, Kentucky is again poised to “steer the ship” as a state seeking innovative educational reform. As educators await the next steps in education, the Commonwealth has an invaluable opportunity for further research and reflection in the field of the SISI and Scholastic Audit process—two measures that could hasten future reform in Kentucky and throughout the nation, such as the Next Generation Science Standards (NGSS).

In the current study, the researcher investigated the impact of the Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom

Evaluation/Assessment, and Standard 3, Instruction—on the elementary science component of the Academic Index, while controlling for demographic control variables. Due to the specificity of the current research, several topics were beyond the scope of this investigation and warrant further investigation. In the section below, several topics are listed for future research.

First, since the SISI—the guiding document of the Scholastic Audit—are general content standards, not specific to science that is the focus of the current study, one wonders if other improvement models (e.g., the Literacy Design Collaborative model, the Danielson Framework, Laying the Foundation Program) that educational reformers propose are set up generally or differentially to address other content areas. Currently, there are no data on this topic. More research is needed to determine if these models would look different if they were subject specific.

Second, since the current research only addresses the Academic Performance standards for the elementary science portion of the Academic Index, more research is warranted into the influence of the Academic Performance standards on middle school and high school science, while controlling for accompanying demographic factors. No current studies exist that investigate the Academic Performance standards at these two levels.

Third, regarding the content area of science addressed by the current research, more research is needed on the influence of the Academic Performance standards on other content areas—social studies, ELA, mathematics—at a variety of grade levels. Through an investigation of the impact of Academic Performance standards on different content areas, educators would be afforded a more robust, comprehensive guide to school reform across all content areas and, thus, a more thorough look into the efficacy of the SISI standards both generally and differentially.

Fourth, since the current study only addresses the Scholastic Audit cycles between 2001-2005, more research is warranted on the influence of the Academic Performance standards on the elementary science portion of the Academic Index during testing cycles beyond 2005. As new accountability structures in the past 11 years have been utilized by the Kentucky Department of Education (KDE), further investigation of these accountability measures, and the subsequent impact of the Academic Performance on these measures, is needed.

Fifth, this is the only study to examine the current sample of elementary schools ($N = 181$) by removing 17 incidences of repeat Scholastic Audits—for a purposeful, adjusted sample of $n = 164$. Therefore, more research is needed on these 17 specific schools to determine what factors contributed to necessitate a repeat Scholastic Audit at these aforementioned schools, during the 2001-2005 time frame.

Sixth, in the current study, Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction) from the SISI document are addressed. Additionally, for the current set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008; Todd, 2010) chaired by Drs. Smith and Miller regarding the SISI document and the Scholastic Audit database, the following combinations of standards were examined: B. C. Ennis (2007) assessed Standards 6 (Professional Development, Professional Growth and Evaluation) and 7 (Leadership) at the elementary level; McKinney (2007) investigated Standard 7, with Standards 1 and 3 as mediating variables, at the elementary level; Saravia (2008) examined Standards 4 (School Culture) and 5 (Student, Family and Community Support) at the elementary level; and, Todd (2010) replicated the research of McKinney (2007) at the high school level. Therefore, since the SISI document is composed of nine standards for whole school reform, more research is

needed for the remaining standards and their individual and collective influence on the elementary science portion of the Academic Index.

Seventh, since the abandonment of the *Standards and Indicators for School Improvement* (SISI) by the Kentucky Department of Education due to concerns about cost, a team of Western Kentucky University researchers have produced a revision and replacement for the SISI termed the *School Improvement Scholastic Review* (SISR). The SISR is based on teacher perceptual scales that expedite the data gathering process through a non-invasive, inexpensive means of data collection. Early piloting of the SISR instrument has yielded results comparable to that of the Scholastic Audit process, anchored by the SISI. Therefore, more research is needed using the *School Improvement Scholastic Review* as a method for whole school reform and improvement.

Eighth, due to the possibility of more aggressive special education testing in the wealthier school districts of Kentucky, with testing modifications being provided to identified students, there is a subsequent possibility that demographic factors related to poverty could be overlooked as a catalyst in giving students special education—as the overall result regarding student achievement masks trends of subsets of socioeconomic status that progress in opposite directions. Reeves and Lowe (2009) stated, “the very success of education policy may hinge upon policymakers acquiring fine-grained knowledge of the varying responses to policy that are likely to occur in the target population” (p. 176). While one answer to addressing this potential trend of special education testing occurring differentially at different levels of the socioeconomic spectrum is employing split samples, another solution to provide a more in-depth examination of these potential trends is quantile regression, which is a type of statistical analysis that employs the median m , noted as “the point within any given distribution where half the

values are greater than m and the remaining values are equal or less than m " (p. 180). Reeves and Lowe added that the use of the median m was much more reliable when dealing with non-normal distributed data, such as income (p. 180). Furthermore, Reeves and Lowe urged education policy researchers to consider quantile analysis as "it can no longer be assumed that conditional mean relations are optimal summaries of the constraints experienced by educational target groups" (p. 193). Quantile analysis could serve as an ideal means of investigating the possibility of more aggressive testing for special education populations in wealthier school districts—elucidating perhaps a variety of trends and implications for how the "pressures of standardized testing," socioeconomic status, and levels of performance expectations influence schools and school districts in their decisions to identify target populations. Therefore, more research is needed to investigate the propensity of special education testing based on the socioeconomic status of the schools and school districts. Additionally, while not as rigorous nor aggressive as the special education identification process, the identification of gifted students serves as another area where quantile testing could serve as grounds for future research.²

While the full implications of the ESSA are not yet known, with the proper foresight and reflective hindsight, Kentucky can again spearhead the reform movement for the nation. In the following section, the researcher concludes the current study through the lens of the central research question.

Conclusions

Out of the darkness and into the light, the landmark ruling by the Kentucky Supreme Court in 1989—in response to a lawsuit brought by 66 school districts who sought equitable funding in education finance—set the educational reform gears in motion for Kentucky and the hundreds of thousands of students within its borders. Capturing the

zeitgeist of reform sweeping the nation, Kentucky spearheaded its own educational improvement measures through the Kentucky Education Reform Act (KERA) of 1990 that prompted then Governor Wallace Wilkinson to proclaim: “History, it has been said, is a river flowing through time. If that is so, then today we are changing its course. On this day, more than any other, I am proud to be a Kentuckian, and I am proud to be your governor” (Steffy, 1993, p. 4).

Through the billions of dollars that funded Kentucky educational reform, the Kentucky Department of Education (KDE) composed a whole school reform document, the *Standards and Indicators for School Reform (SISI)*, that actualized the ideals of KERA through nine standards and 88 respective indicators—a model of whole school improvement that could be accessed by any educator across the Commonwealth. The nine standards of the SISI were divided into three domains of school improvement: Academic Performance, Learning Environment, and Efficiency. Additionally, the SISI served as the guiding document for the Scholastic Audit process, whereby struggling schools were offered specific areas of reform, according to the SISI, by trained teams of educators (and parents) to help guide these schools toward continuous progress and the goal of proficiency by 2014. For the current study, the researcher retrieved information from the KDE-maintained secondary database of Scholastic Audit data regarding the three Academic Performance standards—Standard 1 (Curriculum), Standard 2 (Classroom Evaluation/Assessment), and Standard 3 (Instruction).

The purpose of this study was to determine the influence of the SISI Academic Performance standards—Curriculum, Classroom Evaluation/Assessment, and Instruction—on the elementary science component of the Academic Index and to explore possible relationships among demographic factors such as Percentage Free and Reduced

Lunch, Appalachian Status, and Percentage White. The current study focused only on elementary schools and, specifically, the fourth grade—when students take the elementary science assessment. Other studies in the current set of dissertations (B. C. Ennis, 2007; McKinney, 2007; Saravia, 2008) chaired by Drs. Smith and Miller also examined SISI standards at the elementary level. Additionally, Todd (2010) examined the influence of SISI standards on the Academic Index, but at the high school level. The current research is the only study to investigate all three Academic Performance standards at the elementary school level.

The current study offers the following results: (a) the inclusion of all specified Demographic Factors produce a significant impact (effect size of .626) on the elementary science component of the Academic Index; (b) the three Academic Performance standards collectively account for 28% of the overall variance on the elementary science portion of the Academic Index; (c) Curriculum does not yield a significant influence on the elementary science component of the Academic Index; (d) Classroom Evaluation/Assessment yields a significant impact (effect size of .349) on the elementary science portion of the Academic Index; (e) Instruction comes close to significance ($p = .073$) in influence on the elementary science component of the Academic Index; and (f) the final hierarchical regression model—including only significant demographic factors and Classroom Evaluation/Assessment—accounted for 68% of the overall variance on the elementary Science Academic Index.

After performing psychometric analyses such as Factor Analysis, Cronbach's (1951) coefficient alpha, and scree plots, the researcher determined that the results from the separate calculations offered a one factor solution regarding the respective indicators for each of the three Academic Performance standards. In terms of the influence of

demographic factors on each of the three Academic Performance standards, demographic control variables accounted for very little of the variance for each respective standard. For Standard 1 (Curriculum), demographic factors accounted for 13% of the variance, with only Year of Audit (effect size of .232) and Voluntary versus Involuntary status (effect size of -.206) yielding significant effects. For Standard 2 (Classroom Evaluation/Assessment), demographic control variables explained 16.7% of the overall variance, with only Year of Audit (effect size of .213) producing a significant effect. For Standard 3 (Instruction), demographic factors accounted for 17.3% of the total variance for Instruction, with only Voluntary versus Involuntary status (effect size of -.181) yielding a significant impact. Much like the findings of Saravia (2008) for School Culture and Student, Family and Community Support, the results of the current study are of particular interest to high-poverty schools and school districts, as Academic Performance remains fairly independent of the effects of Demographic Factors. Therefore, schools can confidently employ these alterable strategies inherent in the three Academic Performance standards knowing their efforts are relatively free of demographic influence.

While the three Academic Performance standards accounted for 28% of the variance, only Standard 2 (Classroom Evaluation/Assessment) had a significant impact on the elementary science portion of the Academic Index. Curriculum did not approach significance, while Instruction was almost significant ($p = .073$). Both McKinney (2007) and Todd (2010) found Instruction to be significant in their studies regarding Standard 1 at the elementary and high school levels, respectively, on the Academic Index. Two possible explanations for this discrepancy exist.

First, the possibility of multicollinearity exists between Standards 2 and 3, with a high correlation between the two standards of $r = .811$ at the $p \leq .01$ level. Therefore, the

significance of Standard 3 (Instruction) could be masked by the high correlation to Standard 2 (Classroom Evaluation/Assessment). McKinney (2007) and Todd (2010) did not examine Standard 2; therefore, the researchers identified Standard 3 as having a significant impact on the Academic Index, without the multicollinearity problem created by Standard 2 (Cohen et al., 2003). Second, one could suppose that the specificity of the current study on only the elementary science portion of the Academic Index relegated Instruction to being not significant, as opposed to the studies of McKinney (2007) and Todd (2010) who assessed the impact of Instruction on the *entire* Academic Index. It is beyond the scope of this study to determine the exact causes of such a finding; however, more research is warranted at both the elementary level—in replicating the findings—and the high school level regarding the influence of the Academic Performance standards on science, and other individual content areas.

Additionally, in alignment with the findings of McKinney (2007) and Todd (2010), Curriculum was not significant—as an influencer in the current research of the elementary science portion of the Academic Index. Since the version of the science assessment during the sample collection from 2001-2005 did not include performance tasks, assessing student comprehension of science concepts without first assessing student reading aptitude is difficult. Therefore, to gauge accurately student science performance, performance tasks and other testing methods must first be added whereby reading acumen then becomes less of a barrier to student success.

Regarding the influence on the elementary science portion of the Academic Index, Demographic Factors yielded a significant and large impact, accounting for nearly 63% of the overall variance on the Academic Index. Furthermore, when considering the final hierarchical regression model of significant demographic factors and Classroom

Evaluation/Assessment (the only significant Academic Performance standard), the three SISI Academic Performance standards only added an additional 5 percent to the model. The large influence of Demographic Factors on the Academic Index—specifically the elementary science component for the current research—aligned with the previous work of McKinney (2007), Saravia (2008), and Todd (2010) who found demographic control variables as substantially affecting the entire Academic Index.

There are two sides to this. As previously mentioned, this and related studies do not provide evidence of hope that alterable classroom practices—as evidenced in the influence of the three Academic Performance standards—has much impact on the equity dimension (inequalities in schools) (Riley, 2016). In contrast, these same related studies do suggest that some aspects of KERA generally, in this study on the Academic Performance section of the *Standards and Indicators for School Improvement*, do have a positive influence in the overall level of achievement (excellence dimension). Thus, for Kentucky educators who refuse to accept the idea of demography as destiny (cf. Miller & Moore, 2006; Roeder, 2000b; Todd, 2010), more robust educational changes are needed (see, e.g., the points enumerated in **Implications for Policy** above).

Lessons Learned

To macro-level policy makers. As Carl Sagan once proclaimed, “Extraordinary claims require extraordinary evidence.” The findings of the current research illuminate both the significant progress on the *excellence* dimension of the Academic Index—as indicated by the Year of Audit finding and other overall indicators of the success of the Kentucky Education Reform Act—but that progress on the *equity* dimension has still yet to materialize—as demographic factors still account for more than 63% of the overall variance in the Science Academic Index scores (Miller & Moore, 2006; Roeder 2000a,

2000b; Riley, 2016). Therefore, for macro-level policy makers, a substantial investment in education—enhanced technological infrastructure, equitable resource funding, and adequate professional learning stipends—could help alleviate much of the issues on the equity dimension in even the poorest school districts in Kentucky (Roeder, 2000a, 2000b).

To micro-level policy makers. The findings of the current research do indicate that the three Academic Performance standards—Standard 1, Curriculum, Standard 2, Classroom Evaluation/Assessment, and Standard 3, Instruction—can palpably affect student achievement, in light of—and in spite of—the influence of Demographic Factors. During Step 2 of the hierarchical multiple regression equation, with the introduction of the SISI Academic Performance standards, the influence of both Voluntary versus Involuntary status and Percentage Free and Reduced Lunch were reduced, from $-.286$ to $-.248$ and $-.257$ to $-.229$, respectively. Therefore, sound classroom practices, as indicated by the *Standards and Indicators for School Improvement*, can and do make a difference in the lives of students through enhanced student achievement outcomes.

While the *Standards and Indicators for School Improvement* are no longer applicable, since the Kentucky Department of Education discontinued the Scholastic Audit (and, therefore, the SISI as well), the Western Kentucky University education team has developed a replacement, the *School Improvement Scholastic Review* (SISR), based on teacher perceptual scales—and is thus less expensive, more feasible, and, based on the data to date, equally effective for a school improvement tool (Dr. Stephen Miller, personal communication, July 16, 2016).

Therefore, to the micro-level policy makers, two suggestions are afforded through the current research. First, in alignment with the significant influence of Standard 2, Classroom Evaluation/Assessment, on the Science Academic Index, more professional

learning opportunities that focus on best practices in assessment (Stiggins, 2000; Wiggins, 1990; Zemelman et al., 2012) could bolster current student achievement. These professional learning opportunities could be provided by the Kentucky Department of Education or by regional educational cooperatives (such as the Ohio Valley Educational Cooperative or the Green River Regional Educational Cooperative). Second, through the implementation of an annual or biannual administration of the inexpensive, effectual *School Improvement Scholastic Review* (SISR) tool at every school in the Commonwealth, schools could be afforded a whole school reform model to help educators at every level in the district to approximate those best practices that lead to school improvement.

To administrators. Through the current research, the absence of Standard 3, Instruction, as a significant factor in predicting Science Academic Index scores underscores the importance of the current iteration of the science assessment as more of a reading examination—as students sift through reading passages before answering with science content—than a true measure of science knowledge. Therefore, to the administrators, two suggestions become increasingly apparent in light of the current research.

First, administrators should encourage teachers to incorporate more science-infused nonfiction reading material into the English Language Arts content of the Common Core State Standards. Not only does science content enhance engagement and achievement (Romance & Vitale, 1992), but science also becomes the pathway to curiosity that leads to lifelong learning—a pathway that can lead to enhance measures of college-and-career readiness later. Second, administrators would be well-served to implement stand-alone science classes across grade levels and especially at the elementary level. To achieve such a measure as stand-alone science classes, in light of the ever-increasing burden on teachers

to cover all the requisite skills and standards during the school day, administrators must be willing to think outside of the box in terms of creative scheduling. One solution to allow for stand-alone science classes during the elementary level school day is the employment of a full-time science teacher to address the science needs of the entire building.

Subsequently, this science teacher would thereby meet once a week with each classroom of students during “specials” or “pull-out” time—much like a gym, music, or related arts teacher. Additionally, to alleviate the obstacle of where to house such a teacher in a possibly over-crowded elementary school and whereby a room for a science laboratory is unfeasible, the full-time science teacher could move from classroom to classroom during a respective pull-out block of science instruction.

To education practitioners. With the significance of Standard 2, Classroom Evaluation/Assessment, on the Science Academic Index, best practices in assessment become even more viable as pathways to enhanced learning outcomes for students. Therefore, to the education practitioner, best practices such as formative assessments and assessment for learning should be employed daily in the classroom to provide a continuous channel of student feedback and progress toward the mastery of standards and content (Clarke, 2008). Additionally, in terms of improving current assessment practices, the incorporation of authentic assessments into the classroom represents another invaluable strategy for increasing student mastery and achievement (Wiggins, 1990). Finally, perhaps the last stage of the recursive assessment process, meaningful feedback to students serves as the vehicle to drive learning—and to enhance perseverance as feedback encourages modifications to the current level of thinking or iteration of authentic product. Stiggins (2012) encouraged classroom educators to employ product-focused, actionable, descriptive, and clear feedback to learners (p. 4).

Since the inception of the Kentucky Education Reform Act of 1990, Kentucky educators have witnessed their fair share of innovative educational salves come and go as school reformers searched endlessly for the unifying educational theory, the pedagogical panacea, that would bring Kentucky education to the lofty goal of proficiency by 2014. Two years since the 2014 goal, and with a new Commissioner of Education and federal education act, Kentucky now has its sights set on a new priority: to have students fully prepared to be college- and career-ready. With student achievement still a large measure—and standard indicator—of student success, perhaps Kentucky is finally ready to be honest about the impact of demographic factors—specifically, poverty as indicated through the Percentage Free and Reduced Lunch in the current research—on student success, both inside and outside the classroom. With no time for excuses nor apathy, Kentucky must continue to make a significant and substantial investment in the education system. However, perhaps even more important, Kentucky educators must begin to focus on what they can control and what they cannot—specifically, those alterable practices of Classroom Evaluation/Assessment—that can influence student success in the classroom. Only then will it be possible truly to bring Kentucky education out of the darkness and into the light, where, ultimately, the lives of children in the Commonwealth can expect to achieve both *excellence* and *equity* as they prepare for a life of global challenges.

ENDNOTES

¹ Regarding the potential inclusion of Steps 3 and 4 in the hierarchical multiple regression model, the researcher investigated removing Standards 1 (Curriculum) and 3 (Instruction) of the SISI Academic Performance standards in subsequent steps; however, the increase in β and Adjusted R^2 values did not add enough value to necessitate two additional steps for the hierarchical multiple regression model.

² See DeYoung (1993) for an explanation of the inequities in funding that lead to reduced services being provided for special student populations in rural school districts.

REFERENCES

- 703 Kentucky Administrative Regulations 5:120.
- Ackoff, R. L. (1981). *Creating the corporate future*. New York, NY: John Wiley & Sons.
- Aguiar, O., Mortimer, E., & Scott, P. (2010). Learning from and responding to students' questions: The authoritative and dialogic tension. *Journal of Research in Science Teaching*, 47, 174-193.
- Akerson, V., Cullen, T., & Hanson, D. (2009). Fostering a community of practice through a professional development program to improve elementary teachers' views of the nature of science and teaching practice. *Journal of Research in Science Teaching*, 46, 1090-1113.
- Alston, E., Brinly, B., Carr, A., Deaton, S., Dutton, P., Little, D., & Steinburg, E. (1999). *Kentucky Education Reform Act: A citizen's handbook*. Frankfort, KY: Legislative Research Commission.
- Appalachia Educational Laboratory. (1995). *The needs of Kentucky teachers for designing curricula based on academic expectations*. Charleston, WV: Author.
- Appalachia Educational Laboratory. (2002). *Findings of the Kentucky Scholastic Audit/Review survey*. Retrieved from <http://www.earchives.ky.gov/pubs/Education/Research%20reports/AELAuditFinalReport103102.pdf>
- ASCD. (2015). *ESEA reauthorization updates*. Retrieved from <http://www.ascd.org/public-policy/esea-reauthorization-updates.aspx>

- Azzam, A. M., Perkins-Gough, D., & Thiers, N. (2006). The impact of NCLB. *Educational Leadership, 64*, 94-96.
- Baeder, J. (2016) *Why everyone believes in formative assessment but nobody does it*. Retrieved from <http://www.principalcenter.com/why-everyone-believes-in-formative-assessment-but-nobody-does-it/>
- Berger, R., Rugen, L., & Woodfin, L. (2014). *Leaders of their own learning: Transforming schools through student-engaged assessment*. San Francisco, CA: Jossey-Bass.
- Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Retrieved from http://www.spd.dcu.ie/site/teaching_today/documents/raisingstandardsthroughclassroomassessment.pdf
- Bloom, B. S. (1980). The new direction in educational research: Alterable variables. *Phi Delta Kappan, 61*, 382-385.
- Bowles, S. A., Churchill, A. M., Effrat, A., & McDermott, K. A. (2002). *School and district intervention: A decision-making framework for policymakers*. Amherst, MA: University of Massachusetts Center for Education Policy.
- Bray, B., & McClaskey, K. (2015). *Make learning personal: The what, who, wow, where, and why*. Thousand Oaks, CA: Corwin.
- Brown, R., Brown, J., Reardon, K., & Merrill, C. (2011). Understanding STEM: Current perceptions. *The Technology and Engineering Teacher, 70*, 5-9.
- Bybee, R. W. (Ed.). (2002). *Learning science and the science of learning*. Arlington, VA: NSTApress.
- Campbell, J., Voelkl, K., & Donahue, P. (1996). *NAEP 1996 trends in academic progress*. Washington, DC: National Library of Education, Office of Educational Research

and Improvement.

- Capra, F. (1997). *The web of life*. New York, NY: Doubleday-Anchor Book.
- Caudill, H. M. (1963). *Night comes to the Cumberlands: A biography of a depressed area*. Boston, MA: Little, Brown.
- Checkland, P. (1997). *Systems thinking, systems practice*. Chichester, England: John Wiley & Sons Ltd.
- Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: Opportunities and implementation. *Shanghai Archives of Psychiatry*, 26, 371-375.
- Civic Impulse. (2015). *S. 1513 – 103rd Congress: Improving America's School Act of 1994*. Retrieved from <https://www.govtrack.us/congress/bills/103/s1513>
- Clarke, S. (2008). *Active learning through formative assessment*. London, England: Hodder Education.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10, 1-9.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle: University of Washington, Center for the Study of

Teaching and Policy.

Darling-Hammond, L. (2007). A Marshall plan for teaching. *Education Week*.

Retrieved from

<http://www.edweek.org/ew/articles/2007/01/10/18hammond.h26.html>

DeSilver, D. (2015). *U.S. students improving—slowly—in math and science, but still lagging internationally*. Pew Research Center. Retrieved from

<http://www.pewresearch.org/fact-tank/2015/02/02/u-s-students-improving-slowly-in-math-and-science-but-still-lagging-internationally/>

“Development Overview.” (2016). Retrieved from

<http://www.nextgenscience.org/development-overview>

DeYoung, A. J. (1993). *Children at risk in America's rural schools: Economic and cultural dimensions*. Retrieved from <http://files.eric.ed.gov/fulltext/ED362339.pdf>

DuFour, R., DuFour, R., Eaker, R., & Many, T. (2006). *Learning by doing: A handbook for Professional Learning Communities at work*. Bloomington, IN: Solution Tree.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences*, 8, 391–450.

Ennis, B. C. (2007). *Leadership, professional development, and accountability scores: Evidence from Kentucky Scholastic Audits* (Unpublished doctoral dissertation).

University of Louisville, Louisville, KY.

Ennis, L. S. (2002). *Effects of instructional strategies on seventh grade science achievement as perceived by Kentucky students* (Unpublished doctoral dissertation). University of Louisville, Louisville, KY.

Evans, R. (2005). Reframing the achievement gap. *Phi Delta Kappan*, 86, 582-589.

- Evers, W. M., & Walberg, H. J. (2004). *Testing student learning, evaluating teaching effectiveness*. Stanford, CA: Hoover Institution Press.
- Fastmetrics. (2015). *Internet connection speed map: State vs state*. Retrieved from <https://www.fastmetrics.com/internet-connection-speed-map-usa.php#average-speeds-by-state>
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28, 465-498.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE Publications Inc.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286-299.
- Forbes, C., & Davis, E. (2009). Beginning elementary teachers' beliefs about the use of anchoring questions in science: A longitudinal study. *Science Education*, 94, 365-387.
- Fram, M. S., Miller-Cribbs, J. E., & Van Horn, L. (2007). Poverty, race, and the contexts of achievement: Examining educational experiences of children in the U.S. South. *Social Work*, 52, 309-319.
- Fulp, S. (2002). *2000 National survey of science and mathematics education: Status of elementary school science teaching*. Retrieved from http://2000survey.horizon-research.com/reports/elem_science/elem_science.pdf
- Fuhrman, S. H. (1999). The new accountability. *CPRE Policy Briefs*. New Brunswick, NJ: Consortium for Policy Research in Education.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

- Geier, R., Blumenfeld, P., Marx, R., Krajcik, J., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45, 922-939.
- Glickman, C., Gordon, S., & Ross-Gordon, J. (2007). *Supervision and instructional leadership* (7th ed.). Boston, MA: Allyn and Bacon.
- Gorard, S. (2002). The role of secondary data in combining methodological approaches. *Educational Review*, 54, 231-237.
- Harms, N., & Yager, R. (1981). *What research says to the science teacher* (Vol. 3). Washington, DC: National Science Teachers Association.
- Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher*, 39, 120-131.
- Haselton, W. B. (2004). *School finance reform in Kentucky: Impact on student outcomes* (Unpublished doctoral dissertation). University of Louisville: Louisville, KY.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
- Havice, W. (2009). The power and promise of a STEM education: Thriving in a complex technological world. In ITEEA (Ed.), *The overlooked STEM imperatives: Technology and engineering* (pp. 10-17). Reston, VA: ITEEA.
- Hess, F. M. (2005). Commentary: Accountability policy and scholarly research. *Educational Measurement: Issues and Practice*, 24, 53-57.
- Hightower, A. M., Delgado, R. C., Lloyd, S. C., Wittenstein, R., Sellers, K., & Swanson,

- C. B. (2011). *Improving student learning by supporting quality teaching*. Retrieved from http://www.edweek.org/media/eperc_qualityteaching_12.11.pdf
- Hinde, E. T. (2005). Revisiting curriculum integration: A fresh look at an old idea. *The Social Studies*, 96, 105-111.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin Company.
- Huck, S. W. (2000). *Reading statistics and research* (3rd ed.). New York, NY: Addison Wesley Longman.
- Innes, R. G. (2010). *KERA (1990-2010): What have we learned*. Bowling Green, KY: Bluegrass Institute.
- Jackson, M. (2003). *Systems thinking: Creative holism for managers*. Chichester, England: John Wiley & Sons Ltd.
- Kelley, C. Milanowski, A., & Heneman, H. G., III. (1998, April). *Changing teacher compensation: Cross-site analysis of the effects of school-based performance award programs*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Kentucky Association of School Councils. (2012). *Standards and Indicators for School Improvement*. Retrieved from <http://www.kasc.net/2010/Standards%20and%20Indicators%20for%20School%20Improvement.pdf>
- Kentucky Department of Education. (1990). *Kentucky Education Reform Act of 1990*. Retrieved from http://cber.uky.edu/Downloads/kentucky_education_reform_act.htm
- Kentucky Department of Education. (1993). *Transformations: Kentucky's Curriculum*

- Framework* (Vols. 1-2). Frankfort, KY: Author.
- Kentucky Department of Education. (1996). *Core content for assessment (version 1.0)*. Frankfort, KY: Author.
- Kentucky Department of Education. (1998). *Program of studies (version 1.5)*. *Implementation manuals* [CD-ROM]. Frankfort, KY: Author.
- Kentucky Department of Education. (2000a). *Core content for science assessment, version 3.0*. Frankfort, KY: Author.
- Kentucky Department of Education. (2000b). *Results matter: A decade of difference in Kentucky's public schools*. Frankfort, KY: Author.
- Kentucky Department of Education. (2003). *The Scholastic Audit: A report on school improvement in Kentucky*. Frankfort, KY: Author.
- Kentucky Department of Education. (2004a). *CATS interpretive guide*. Retrieved from <http://www.education.ky.gov>
- Kentucky Department of Education. (2004b). *School level performance descriptors for Kentucky's Standards and Indicators for School Improvement*. Frankfort, KY: Author.
- Kentucky Department of Education. (2004c). *Standards and Indicators for School Improvement: A Kentucky model for student-centered accountability*. Frankfort, KY: Author.
- Kentucky Department of Education. (2005a). *Kentucky's core content for assessment, Version 4.0*. PowerPoint presented by S. Lewis and N. LaCount for ISN. Retrieved from http://www.powershow.com/view/3c41-ZTA4M/Kentuckys_Core_Content_for_Assessment_Version_4_O_powerpoint_pp_t_presentation

Kentucky Department of Education. (2005b). *Understand the SISI* [PowerPoint slides].

Retrieved from email communication with R. Hurley, October 21, 2015.

Kentucky Department of Education. (2006a). *CATS interpretive guide*. Retrieved from

<http://www.education.ky.gov>

Kentucky Department of Education. (2006b). *Core content for science assessment, version*

4.1. Retrieved from <http://www.education.ky.gov>

Kentucky Department of Education. (2006c). *Program of studies*. Retrieved from

<http://www.education.ky.gov>

Kentucky Department of Education. (2008). *School level performance descriptors and*

glossary for Kentucky's Standards and Indicators for School Improvement.

Retrieved from <http://education.ky.gov>

Kentucky Department of Education. (2012a). *College and career readiness unified plan*.

Retrieved from <http://education.ky.gov>

Kentucky Department of Education. (2012b). *FAQs on Kentucky's new assessment and*

accountability system for public schools. Retrieved from <http://education.ky.gov>

Kentucky Department of Education. (2012c). *Kentucky's learning goals and academic*

expectations. Retrieved from

[http://education.ky.gov/curriculum/docs/Documents/Kentucky's%20Learning%20Goals%20and%20Academic%20Expectations.pdf](http://education.ky.gov/curriculum/docs/Documents/Kentucky%20Learning%20Goals%20and%20Academic%20Expectations.pdf)

Kentucky Department of Education. (2013). *Kentucky education facts*. Retrieved from

<http://education.ky.gov/comm/edfacts/Pages/default.aspx>

Kentucky Department of Education. (2015a). *K-Prep*. Retrieved from

<http://education.ky.gov/aa/assessments/pages/k-prep.aspx>

Kentucky Department of Education. (2015b). *Kentucky's learning goals and academic*

- expectations*. Retrieved from
<http://education.ky.gov/curriculum/standards/kyacadstand/Pages/Kentucky's-Learning-Goals-and-Academic-Expectations.aspx>
- Kentucky Revised Statute 158.6455. Retrieved from
<http://www.lrc.ky.gov/Statutes/statute.aspx?id=3556>
- Knight, J. (2011). What good coaches do. *Coaching: The New Leadership Skill*, 69, 18-22.
- Koger, M. E., & Thacker, A. A. (2004). *Comparative study of Standards and Indicators of School Improvement (SISI) and Academic Index for selected elementary schools*. Louisville, KY: Human Resources Research Organization (HumRRO).
- Kress, S., Zechmann, S., & Schmitt, J. M. (2011). When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Journal on Legislation*, 48, 185-234.
- Lasky, S. G. (2004). *Toward a policy framework for analyzing educational system effects* (Report No. 71). Center for Research on the Education of Students Placed at Risk.
- Laszlo, A., & Krippner, S. (1998). Systems theories: Their origins, foundations, and development. In J. S. Jordan (Ed.). *Systems theories and a priori aspects of perception* (pp. 47-74). Amsterdam, Holland: Elsevier Science.
- Lee, H., Linn, M., Varma, K., & Liu, O. (2010). How do technology-enhanced inquiry science units impact classroom learning. *Journal of Research in Science Teaching*, 47, 71-90.
- Legislative Research Commission. (2005). *An analysis of the Commonwealth Accountability Testing System*. Frankfort, KY: Legislative Research Commission. Retrieved from <http://www.lrc.ky.gov/lrcpubs/rr328.pdf>
- Legislative Research Commission. (2006). *Highly skilled educator program*. Frankfort,

- KY: Legislative Research Commission. Retrieved from <http://www.lrc.ky.gov/lrcpubs/rr339.pdf>
- Legislative Research Commission. (2013). *Test familiarity and performance: Comparing scores on Kentucky Core Content Tests and unfamiliar tests*. Frankfort, KY: Legislative Research Commission. Retrieved from <http://www.lrc.ky.gov/lrcpubs/rr402.pdf>
- Levin, H. M. (1974). A conceptual framework for accountability in education. *The School Review*, 82, 363-391.
- Lindle, J. C. (2001). Curriculum reform, the Educational Professional Standards Board, and the Kentucky Department of Education. In J. M. Petrosko & J. C. Lindle (Eds.). *2000 review of research on the Kentucky Education Reform Act* (pp. 83-114). Lexington: University of Kentucky/University of Louisville Joint Center for the Study of Educational Policy and Kentucky Institute for Education Research.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32, 3-13.
- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31, 3-16.
- Lockwood, A. T. (1994). The iron triangle [Interview with Stephen Miller, School effectiveness research]. *Focus in Change*, 16, 15-17.
- Lyons, R., & Barnett, D. (2011). School audits and school improvement: Exploring the variance point concept in Kentucky's elementary, middle, and high schools. *International Journal of Education Policy and Leadership*, 6, 1-15.

- Magaña, S., & Marzano, R. J. (2014). *Enhancing the art & science of teaching with technology*. Bloomington, IN: Marzano research Laboratory.
- Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us*. Aurora, CO: Mid-continent Research for Education and Learning.
- Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2013). *Becoming a high reliability school: The next step in school reform*. Bloomington, IN: Marzano Research Laboratory.
- Marzano, R. J., & Brown, J. L. (2009). *A handbook for the art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., & Marzano, J. S. (2003). The keys to classroom management. *Educational Leadership*, 61, 6-13.
- McKinney, L. E. (2007). *Leadership, curriculum, instruction, and accountability scores: Evidence from Kentucky Scholastic Audits* (Unpublished doctoral dissertation). University of Louisville, Louisville, KY.
- McKinney, L. E., Miller, S. K., & Smith, D. C. (2007, November). *Psychometric analysis of the curriculum, instruction, and leadership standards from Kentucky's Standards and Indicators for School Improvement*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Hot Springs, AR.
- Mele, C., Pels, J., & Polese, F. (2010). A brief review of systems theories and their managerial applications. *Service Science*, 2, 126-135.
- Miller, S. K. (1992). Changing conceptions of "good schools": Implications for reforming urban education. *Education and Urban Society*, 25, 71-84.
- Miller, S. K., Chon, K. H., Houchens, G. W., & Hunt, R. (2013). *Revising Kentucky's*

- Standards and Indicators for School Improvement: Rationale, conceptualization, and resulting instrument.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Pensacola Beach, FL.
- Miller, S. K., Houchens, G. W., Smith, D. C., Chon, K. H., & Hunt, R. (2014a, April). *Kentucky's Standards and Indicators for School Improvement* (Unpublished manuscript). Western Kentucky University, Bowling Green, KY.
- Miller, S. K., Houchens, G. W., Smith, D. C., Chon, K. H., & Hunt, R. (2014b, May). *Replacing the Scholastic Audit with teacher perceptual scales: Psychometric analysis of the Standards and Indicators Scholastic Review.* Paper presented at the International Conference on Learning and Higher Education, Nashville, TN.
- Miller, S. K., & Moore, B. D. (2006). *Whither race? Demographic and grade level influences on reading assessment in a desegregated metropolitan district.* Unpublished manuscript, University of Louisville, Louisville, KY.
- Miller, S. K., Smith, D. C., & Ennis, L. S. (2006). The effects of race, place, class, and gender on instructional strategies in Kentucky's seventh grade science classes: Individual and school level analyses. *Southern Rural Sociology, 21*(2), 65-88.
- Mintrop, H. (2003). The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives, 11*(3).
- Mintrop, H., & Trujillo, T. (2005). Corrective action in low performing schools: Lessons for NCLB implementation from first-generation accountability systems. *Education Policy Analysis Archives, 13*(48).
- Morrison, J., Raab, F., & Ingram, D. (2009). Factors influencing elementary and secondary teachers' views on the nature of science. *Journal of Research in Science*

Teaching , 46, 384-403.

Murphy, J. (2004). Leadership for literacy: A framework for policy and practice. *School Effectiveness and School Improvement*, 15, 65-96.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the Nation and the Secretary of Education*. Washington, DC: Superintendent of Documents, Government Printing Office.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (J. D. Bransford, A. L. Brown, & R. R. Cocking, Eds.). Washington, DC: National Academy Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Committee on Science learning, Kindergarten Through Eighth Grade (R. A. Duschl, H. A. Schweingruber, & A. W. Shouse, Eds.). Board on Science Education, Center for Education. Division of Behavioral and Social Science and Education. Washington, DC: The National Academies Press.

National Research Council. (2010). 4.0: The ingredients of innovation. In Members of the 2005 "Rising Above the Gathering Storm" Committee (Eds.) *Rising above the gathering storm, revisited: Rapidly approaching category 5* (pp. 43-62). Washington, DC: National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

National Science Teachers Association. (2013). *NSTA position statement: The next generation science standards*. Retrieved from <http://www.nsta.org/about/positions/ngss.aspx>

- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
- Ogawa, R. T., Sandholtz, J. H., Martinez-Flores, M., & Scribner, S. P. (2003). The substantive and symbolic consequences of a district's standards-based curriculum. *American Educational Research Journal, 40*, 147-176.
- Palmer, D. (2009). Student interest generated during an inquiry skills lesson. *Journal of Research in Science Teaching, 46*, 147-165.
- Pankratz, R. S., & Petrosko, J. M. (Eds.). (2000). *All children can learn: Lessons from the Kentucky reform experience*. San Francisco, CA: Jossey-Bass.
- Petrosko, J. M. (2000). Assessment and accountability. In J. M. Petrosko & J. C. Lindle (Eds.), *2000 review of research on the Kentucky Education Reform Act* (pp. 3-81). Lexington: University of Kentucky/University of Louisville Joint Center for the Study of Educational Policy and Kentucky Institute for Education Research.
- Petrosko, J. M., Lindle, J. C., & Pankratz, R. (2000). Executive summary. In J. M. Petrosko & J. C. Lindle (Eds.), *2000 review of research on the Kentucky Education Reform Act* (pp. 3-81). Lexington: University of Kentucky/University of Louisville Joint Center for the Study of Educational Policy and Kentucky Institute for Education Research.
- Porter-Magee, K. (2004). Teacher quality, controversy, and NCLB. *The Clearing House, 78*, 26-29.
- Redding, S., & Walberg, H. (2008). *Handbook on statewide systems of support*. Charlotte, NC: Information Age Publishing.

- Reedy, S. (2010). *KERA information*. University of Kentucky. Retrieved from <http://education.uky.edu/site/KERAinformation>
- Reeves, E. B., & Lowe, J. (2009) Quantile regression: An education policy research tool. *Southern Rural Sociology*, 24(1), 175-199.
- Riley, S. P. (2016). *Curriculum, classroom assessment, instruction, and science outcomes: Evidence from Kentucky's Scholastic Audits* (Unpublished doctoral dissertation). University of Louisville, Louisville, KY.
- Roberts, A. (2012). A justification for STEM education. *The Technology and Engineering Teacher*, 71 (online extension). Retrieved from <http://www.iteea.org/mbrsonly/Library/TTT/TTTe/04-12roberts.pdf>
- Roeder, P. W. (2000a). *Education reform and equitable excellence: The Kentucky experiment (Update-June 2000), Race, poverty, and urban schools in Kentucky*. Retrieved from <http://marces.org/mdarch/pdf/1000021.pdf>
- Roeder, P. W. (2000b). Education reform and equitable excellence: The Kentucky report. *Foresight*, 7, 1-5.
- Romance, N. R., & Vitale, M. R. A curriculum strategy that expands time for in-depth elementary science instruction by using science-based reading strategies: Effects of a year-long study in grade four. *Journal of Research in Science Teaching*, 29, 545-554.
- Rothstein, R. (2010). *How to fix our schools*. Economic Policy Institute. Retrieved from <http://www.epi.org/publication/ib286/>
- Sandall, B. R. (2003). Elementary science: Where are we now? *Journal of Elementary Science Education*, 15, 13-30.
- Saravia, A. (2008). *Elementary parent involvement, school culture, and accountability*

- scores: *Data from Kentucky's Scholastic Audits* (Unpublished doctoral dissertation). University of Louisville, Louisville, KY.
- Sautter, J. M. (2014). Secondary analysis of existing data in social science capstone research. *Council on Undergraduate Research, 34*, 24-30.
- Senge, P. M. (2006). *The fifth discipline: The art & practice of the learning organization*. New York, NY: Doubleday.
- Shymansky, J. A., Kyle, W. C., & Alport, J. M. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching, 20*, 387-404.
- Slavin, R. E. (2007). *Educational research in an age of accountability*. Boston, MA: Pearson Education, Inc.
- Smith, D. C. (2005). Appalachian and rural student performance differences on Kentucky's educational assessment: Fourth grade results. *Research in the Schools, 12*(1), 63-73.
- Smith, D. C. (2006). Appalachian and rural student performance differences on Kentucky's educational assessment: 8th grade results. *Southern Rural Sociology, 21*(2), 47-64.
- Smith, D. C., Ennis, B. C., Saravia, A., Miller, S. K., & Wagner, C. R. (2008, November). *Psychometric analysis of the Learning Environment standards from Kentucky's Standards and Indicators for School Improvement*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN.
- Smith, D. C., Harvey, M. J., Hammock, J., & Miller, S. K. (2009, November). *Psychometric analysis of standards 2, 8, and 9 from Kentucky's Standards and Indicators for School Improvement*. Paper presented at the 2009 National Evaluation Institute, Consortium for Research on Educational Accountability and

- Teacher Evaluation (CREATE), Louisville, KY.
- Smith, D. C., & Miller, S. K. (2013). *Rural demographics and the Standards and Indicators for School Improvement: Influence on elementary school accountability in Kentucky*. Invited paper presented at the National Conference on Rural Education Research, Omaha, NE.
- Staver, J., & Bay, M. (1987). Analysis of the project synthesis goal cluster orientation and inquiry emphasis of elementary science textbooks. *Journal of Research in Science Education, 24*, 629-643.
- Stecher, B., Rahn, M., Ruby, A., Alt, M., Robyn, A., & Ward, B. (1997). *Using alternative assessments in vocational education*. Santa Monica, CA: Rand Corporation.
- Retrieved from http://www.rand.org/pubs/monograph_reports/MR836.html
- Steffy, B. (1993). *The Kentucky education reform: Lessons for America*. Lancaster, PA: Technomic Publishing Co.
- Stiggins, R. (2000). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan, 83*, 758-765.
- Stiggins, R. (2008, September). *Assessment FOR learning, the achievement gap, and truly effective schools*. Paper presented at the 2008 Education Testing Service and College Board Conference, Educational Testing in America: State Assessments, Achievement Gaps, National Policy and Innovations, Washington, DC.
- Stiggins, R. (2012, October). *Summary: 7 essential assessment actions – Rick Stiggins presentation*. Sponsored by L. Cass ISD through the Formative Assessment Series, Michigan Assessment Consortium. Retrieved from <http://michiganassessmentconsortium.org/sites/default/files/StigginsSummary-Lewis-CassFASeries.pdf>

- Stronge, J., Ward, T., & Grant, L. (2011). What makes good teachers good: A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education, 62*, 339-355.
- Todd, R. C. (2010). *Leadership, curriculum, instruction, and accountability scores: Evidence from Kentucky Scholastic Audits* (Unpublished doctoral dissertation). University of Louisville, Louisville, KY.
- U.S. Department of Education. (1965). *Elementary and Secondary Education Act of 1965*. Retrieved from <https://sites.google.com/site/publicedforparents/original-text-of-1965-esea-act>
- U.S. Department of Education. (1994). *Improving America's School Act of 1994*. Retrieved from <http://www2.ed.gov/legislation/ESEA/toc.html>
- U.S. Department of Education. (2001a). *No Child Left Behind Act of 2001*. Retrieved from <http://www2.ed.gov/nclb/landing.jhtml>
- U.S. Department of Education. (2001b). *No Child Left Behind Act of 2001, Executive summary*. Retrieved from <http://www.ed.gov/nclb/overview/intro/execsumm.html>
- U.S. Department of Education. (2004). *No Child Left Behind: A toolkit for teachers*. Retrieved from <http://www2.ed.gov/teachers/nclbguide/nclb-teachers-toolkit.pdf>
- U.S. Department of Education. (2013). *Science, technology, engineering, and math: Education for global leadership*. Retrieved from <http://www.ed.gov/stem>
- U.S. Department of Education. (2015a). *Homeroom: The official blog of the U.S. Department of Education (What is ESEA?)*. Retrieved from <http://blog.ed.gov/2015/04/what-is-esea/>
- U.S. Department of Education. (2015b). *Elementary and secondary education act*. Retrieved from <http://www.ed.gov/esea>

- U.S. Department of Education. (2015c). *Every Student Succeeds Act (ESSA)*. Retrieved from <http://www.ed.gov/essa?src=rn>
- Vierra, A., Pollock, J., & Golez, F. (1998). *Reading educational research* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Vogt, W. P. (1999). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Weinberg, G. M. (2001). *An introduction to general systems thinking* (Silver Anniversary ed.). New York, NY: Dorset House Publishing Company.
- Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). *Report of the 2000 national survey of science and mathematics education*. Retrieved from <http://2000survey.horizon-research.com/reports/status/complete>
- White, S. H. (2011). *Beyond the numbers: Making data work for teachers & school leaders* (2nd ed.). Englewood, CO: The Leadership and Learning Center.
- Wiggins, G. (1990). *The case for authentic assessment*. Retrieved from ERIC database. (ED328611)
- Wilson, C., Taylor, J., Kowalski, S., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, 47, 276-301.
- Zemelman, S., Daniels, H., & Hyde, A. (2012). *Best practice: Bringing standards to life in America's classrooms* (4th ed.). Portsmouth, NH: Heinemann.

APPENDIX A

STANDARDS AND INDICATORS FOR SCHOOL IMPROVEMENT

Appendix A

Standards and Indicators for School Improvement

The *Standards and Indicators for School Improvement* (SISI) document reflected a monumental shift in Kentucky education reform, affording educators a touchstone document through which to gauge and guide whole school improvement efforts for K-12 schools throughout Kentucky (KDE, 2004b). The SISI was composed of nine standards that provided a comprehensive framework for school reform. The SISI standards were further subdivided into three domains of school improvement, each with three consecutive standards: Academic Performance, Learning Environment, and Efficiency. The SISI document included 88 indicators, with each standard comprised of five to 16 indicators, to assist educators in assessing school artifacts and evidence for current implementation of each of the nine standards. The complete list of the 88 indicators for the SISI is included in *School Level Performance Descriptors for Kentucky's Standards and Indicators for School Improvement* (KDE, 2004b).

Academic Performance

Standard 1 (Curriculum) – The school develops and implements a curriculum that is rigorous, intentional, and aligned to state and local standards.

Standard 2 (Classroom Evaluation/Assessment) – The school utilizes multiple evaluation and assessment strategies to continuously monitor and modify instruction to meet student needs and support proficient student work.

Standard 3 (Instruction) – The school's instructional program actively engages all students by using effective, varied, and research-based practices to improve student academic performance.

Learning Environment

Standard 4 (School Culture) – The school/district functions as an effective learning community and supports a climate conducive to performance excellence.

Standard 5 (Student, Family and Community Support) – The school/district works with families and community groups to remove barriers to learning in an effort to meet the intellectual, social, career, and developmental needs of students.

Standard 6 (Professional Growth, Development and Evaluation) – The school/district provides research-based, results driven professional development opportunities for staff and implements performance evaluation procedures in order to improve teaching and learning.

Efficiency

Standard 7 (Leadership) – School/district instructional decisions focus on support for teaching and learning, organizational direction, high performance expectations, creating a learning culture, and developing leadership capacity.

Standard 8 (Organizational Structure and Resources) – There is evidence that the school is organized to maximize use of all available resources to support high student and staff performance.

Standard 9 (Comprehensive and Effective Planning) – The school/district develops, implements, and evaluates a comprehensive school improvement plan that communicates a clear purpose, direction and action plan focused on teaching and learning. (KDE, 2004b)

APPENDIX B

HUMAN SUBJECTS APPROVAL

Appendix B

Human Subjects Approval

**UNIVERSITY OF
LOUISVILLE**

Human Subjects Protection Program Office
MedCenter One – Suite 200
501 E. Broadway
Louisville, KY 40202-1798
Office: 502.852.5188 Fax: 502.852.2164

DATE: November 16, 2015

TO: Thomas R Tretter

FROM: The University of Louisville Institutional Review Board

IRB NUMBER: 15.1118

STUDY TITLE: **Academic Performance, Accountability, and Science Outcomes: Effects of Academic Performance on Elementary Science in Kentucky**

REFERENCE #: 446761

DATE OF REVIEW: 11/12/2015

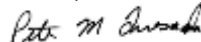
IRB STAFF CONTACT: Name: Sherry Block
Phone: 852-2163
Email: slbloc04@louisville.edu

The IRB Chair/Vice-Chair has reviewed your submission and the project described does not meet the “Common Rule” definition of human subjects’ research. Therefore, this research project does not require IRB review prior to conducting the research.

If you have any questions, please contact the HSPPO office at (502) 852-5188.

Thank you for your submission.

Sincerely,



Peter M. Quesada, Ph.D., Chair
Social/Behavioral/Educational Institutional Review Board

 RE: WKU / UofL Cooperative Doctoral Program Research



Mooney, Paul <paul.mooney@wku.edu>

Tuesday, April 26, 2016 at 8:19 PM

To: Riley, Patrick - (GRREC-New)

Cc: Miller, Stephen; Smith, Douglas

WKU will acknowledge and support the UofL decision.

Paul Mooney
via phone

----- Original message -----

From: "Riley, Patrick - (GRREC-New)" <patrick.riley@grrec.org>

Date: 4/26/2016 8:15 PM (GMT-06:00)

To: "Mooney, Paul" <paul.mooney@wku.edu>

Cc: "Miller, Stephen" <steve.miller@wku.edu>, "Smith, Douglas" <douglas.smith@wku.edu>

Subject: Re: WKU / UofL Cooperative Doctoral Program Research

Professor Mooney,

Please see attached for my Initial Review Submission Packet (Study Application) for the UofL IRB and the Outcome Letter from the UofL IRB. Please let me know if you need any additional documentation.

Thank you for the prompt reply.

Thanks,
Steven Patrick Riley

CURRICULUM VITAE

NAME: Steven Patrick Riley

ADDRESS: 1828 Freeman Avenue
Owensboro, KY 42301

EDUCATION: B.A., Dramatic Arts and English
Centre College
2002

Master of Arts in Teaching, Elementary Education
Oakland City University
2005

Rank I, K12 Principal Certification
Western Kentucky University
2007

Gifted Education Certification, Elementary Education
Murray State University
2012

Doctoral Candidate, P12 Administrative Leadership
University of Louisville & Western Kentucky University
Pending Dissertation Completion

AWARDS: Kentucky Colonel

Leadership Kentucky Alumnus

PRESENTATIONS &
PROFESSIONAL
SERVICE: Kentucky Science Teachers Association (*Presenter / Board*)
2009-2015

International Society for Technology in Education
(*Presenter*)
2014-2016

Cultural Proficiency Institute (*Facilitator*)
2015