VITA

Jane A. Bartley

EDUCATION

2013	Indiana State University, Terre Haute, Indiana Ph.D. in Educational Administration
1985	Indiana State University, Terre Haute, Indiana Master of Education
1974	Indiana State University, Terre Haute, Indiana Bachelor of Science

PROFESSIONAL EXPERIENCE

2009 – Present	Evansville Vanderburgh School Corporation, Evansville, Indiana Director of Elementary, Middle Level, and K-8 Schools
o	
2007 – 2009	Evansville Vanderburgh School Corporation, Evansville, Indiana District Intervention Team
1996 – 2007	Scott Elementary School, Evansville, Indiana Principal
1992 – 1996	Perry Heights Middle School, Evansville, Indiana Assistant Principal
1984 – 1992	Evans Middle School, Evansville, Indiana English Language/Arts Teacher Grades 6-8

THE IMPACT OF FIRST-, SECOND-, AND THIRD-GRADE TEACHERS

ON THIRD-GRADE STUDENT ISTEP+ SCORES

A Dissertation

Presented to

The College of Graduate and Professional Studies

Department of Educational Leadership

Indiana State University

Terre Haute, Indiana

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Jane A. Bartley

May 2013

© Jane A. Bartley 2013

Keywords: Teacher quality, student achievement, teacher accountability, primary grades

COMMITTEE MEMBERS

Committee Chair: Todd Whitaker, Ph.D.

Professor of Educational Leadership

Indiana State University

Committee Member: Terry McDaniel, Ph.D.

Assistant Professor of Educational Leadership

Indiana State University

Committee Member: Charles Watson, Ph.D.

Professor of Education and Chairperson School of Education

University of Evansville

ABSTRACT

This study used existing third-grade ISTEP+ data from a cohort of Evansville Vanderburgh School Corporation students to evaluate various non-experimental methods for estimating teacher effects on student test scores. The study considered the impact that first-, second-, and third-grade teachers had on student performance as measured by the Indiana standardized test in third grade by looking for recognizable patterns of success based on teacher assignment in a cohort of 350 students. By making the assumption of randomness in assigning students to teachers and controlling for student transience, demographics, and teacher movement, the variances of mean ISTEP+ scores were examined to determine and quantify differences based on teacher links. Descriptive statistics summarized possible patterns of success based on teacher links for the cohort as a whole and each school individually by grade level. Differences among teachers by grade level were examined by using an ANOVA model. Regression analysis was used to probe patterns of achievement based on teacher combinations as well as the predictability of ISTEP+ scores based on first-, second-, or third-grade teachers.

ACKNOWLEDGMENTS

This dissertation is the result of the contributions of a number of individuals. I acknowledge the contributions of my dissertation chair, Dr. Todd Whitaker, who knew exactly how to keep me motivated and my spirits up. Often we are our worst critics and Dr. Whitaker allayed my inner doubts about my ability to accomplish this endeavor and supported me in learning to become a researcher. The impact of his gentle prodding, consistent positive feedback, and individualized support cannot be overstated.

Secondly, I thank my dissertation committee members, Dr. Terry McDaniel and Dr. Chuck Watson. I thank Dr. McDaniel and previous EVSC superintendent, Dr. Vince Bertram, specifically for working closely together to make this Ph.D. cohort a reality for many of us who thought that our time had passed. Dr. Watson, whom I have known for years, added that personal support as a critical friend. Both challenged me to think critically and deeply, to ask serious questions throughout the process.

I thank Dr. Eric Hampton for assisting me in designing the analysis models and providing me with a fresh perspective when I became mired in the statistics. Dr. Hampton's frequent review of my content assisted me in keeping clarity of purpose. Thanks also go to Judith Barnes who assisted me in crossing all my t's and dotting all my i's.

Finally, I thank Robert Hochgesang who ran the data and subsequently walked me through its intricacies by explaining the conceptual underpinnings of each statistical model and the output that each provided. I am grateful for the time he invested in my project but most especially for the positive and patient way he taught me how to think in a way that was difficult for me. Special thanks go to the M. A. Rooney Foundation for giving Robert the time within his schedule to take on another project and the EVSC Deputy Superintendent, Susan Riley, for inviting me into relationship with Rooney.

I am also deeply indebted to my family and the EVSC Ph.D. cohort that collectively supported my work and cheered me on.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Hundreds of millions of dollars are spent each year researching and providing resources to K-12 education. The investment is well intended, but good intentions are not enough. Important decisions about schooling happen every day, and to assure that those decisions are sound, relying on evidence is much better than relying on emotion or the way it has always been done. No one wants an undereducated populace, no one believes the economies of the 21st century can survive with a mediocre workforce, and parents most notably want the very best for their children. With parents, community members, governing bodies, and educators focused on the same goal and with ample resource allocation and a substantive motivation to improve the system, it would seem that the task to significantly increase student achievement scores across the nation would be possible. The motivation to change the way education is delivered is clear, but empirical evidence that informs those who make the decisions continues to be a critical element that often is missing. During the past 20 years, research about teaching and learning has increased. Moreover, three conclusions can be drawn: teachers are the most important in-school factor in determining student achievement, teacher effectiveness varies widely, and those differences have a long lasting effect on students (Chetty, Friedman, & Rockoff, 2011; Marzano, 2003). This study attempted to add to the body of research that links teacher performance to students and, additionally, provide information about the impact that first-, second-, and third-grade teachers have on a child's subsequent academic achievement. The implications for policymakers, parents, and especially building administrators are strong. The school a parent chooses for his or her child, the grade level a building principal assigns to teachers, and the investment of finite resources among schools are decisions that would benefit from additional data to inform the decision maker.

Creating and maintaining the multifaceted back-and-forth relationship among individuals that is a necessary catalyst for learning is plagued with variables that significantly impact that interaction (Marzano, 2010). Students and teachers with unique backgrounds built upon diverse external influences produce school cultures that are complicated, sensitive, and constantly evolving. For the most part, studies that focus on the contribution of primary teachers tend to rely on observation and subjective criteria that are difficult to quantify. Humanity is complex and the actualization of learning is difficult, one size does not fit all; and it is deeply emotional and messy most of the time. However, this study endeavored to use inferential statistical analysis and existing data to quantify the contribution of first-, second-, and third-grade teachers in student academic achievement.

The concerns of society regarding the creation and transference of knowledge, processes, and skills are no different than those of societies dating as far back as Socrates. The objective is the same. Generally, human beings want to live together in a safe, comfortable, and efficacious community. In order to create a thriving community, it is imperative that knowledge, processes, and skills be shared and refined from one generation to the next. If this progression does not happen, societies fracture, deteriorate, and then die.

In pursuing this goal, educators and other stakeholders find that the devil is in the details. With over 40 years of solid research as a base, it is now generally held that teachers matter and the quality of these teachers varies considerably (Hattie, 2008). In the 1960s, the common attitude and expectation in the classroom was that teachers could only do the best they could with the students they had. Children from disadvantaged backgrounds and with limited support at home were not expected to achieve at the level of middle class or affluent students (Coleman et al., 1966). However, as social scientists continued to build on the research, the results of many studies began to point to the probability that there was more to the teaching and learning equation. Teacher quality and classroom practice were beginning to bubble up as primary predictors of student achievement (Hattie, 2008).

Good (2010) conducted a study describing how teachers interacted differently with students who they perceived to be high versus low achievers. Subsequently, the study was published and others began to take notice. A study such as Good's stoked the conversation about the interaction between student and teacher. Building on this idea, a field of study based on teacher effectiveness emerged during the 1970s. Individuals who were interested in educational issues continued to examine the student's circumstances of birth and inherent ability as strong predictors. They were influential. They were, however, not the most dominant factor (Anderson, 1979; Nye, Konstantopoulous, & Hedges, 2004). Further studies indicated that effective teachers appeared to be successful with students of all achievement levels regardless of the composition and heterogeneity in their classes (Wright, Horn, & Sanders, 1997). Classrooms in which students were making significant gains in skills acquisition and knowledge were being linked to teacher behaviors, mindsets (Dweck, 2000), expectations (Rosenthal & Jacobson, 1968), and student perceptions of teacher quality (Darling-Hammond, 1997). An articulated knowledge base, deliberate practice, and the ability to know what to do and when to do it, dependent upon the context and the needs of the student, were correlated to student achievement (Ericsson, 1991; Marzano, 2010).

Over time, the research demonstrated incontrovertibly that teachers are key components in affecting student achievement. With a solid foundation to support this premise, the focus began to shift from Coleman's et al.'s (1966) theory of home life and circumstances of birth to quality of classroom experience. Consequently, considerable finger-pointing manifested followed by a myriad of silver bullets designed to fix the problem. Throughout this disorderly period of second order change (Marzano, McNulty, & Waters, 2003), federal and state governments became inexorably linked to schooling. More than ever before, legislators who held the public trusts took that contract seriously and began insisting on accountability or return on investment. Teacher unions, similarly dedicated to protecting the members from assuming all of the responsibility for student learning, dug in. As with any important issue, the narrative swings widely from one position to the other and this movement continues with ferocity today. Eventually the actions tied to the debate will settle upon a centralizing position within this continuum of dissonance, but until that time, children continue to go to school, growing older with each passing day. There is no calling a sidebar to discuss options or opportunities while struggling through emotions and logistics tied to the issue.

Children are born with a natural quest to *know.* Piaget (1954) argued that children sought equilibrium between their mental organizations or schemas and their external environments. Their time is spent being curious and exploring the world, as little children are immersed in an environment of stimuli. With the multitudes of stimuli in today's world of emergent technology, one can surmise that making sense of all that is around them is a full-time job. Deci and Ryan (1985) contended that human beings have an "inherent tendency to seek out novelty and challenges, to extend and exercise their capacities, to explore, and to learn" (p. 27). Novak and Cañas (2002) indicated that preschool children are marvelously adept at meaningful learning; however, upon entering school structured with an emphasis on rote memorization and recall, cognitive connections decrease. Older children are less likely to invest their time on things that they find irrelevant or do not tend to fit with their schema (Kashdan, 2009). They have developed filtering systems that keep out or slough off stimuli that either cause them harm or do not interest them (Sprenger, 2005; Willis, 2010). Schlechty (2002) considered students to be customers, arguing that students make a choice to be involved in their schooling. If the teacher does not deliver novelty, choice, and content that are interesting, the customer will leave and find something elsewhere. The student does not need to get up and physically walk out, he or she just tunes out. Csikszentmihalyi (1990), researcher and philosopher, wrote, "Today's teenagers show the symptoms of the malaise" (p. 15). Learning becomes an external imposition when schooling starts. However, children in the primary grades tend to believe the teacher is all knowing, lives at the school, and does not engage in everyday societal tasks such as grocery shopping. They are inclined to trust in the teacher's competence and character, his or her fundamental goodness and best intentions. Generally primary-grade children do not make conscious decisions regarding the relevance of the lesson or the teacher's instruction (Bronson & Merryman, 2010).

The young child's less discriminating attitudes about what is important to learn afford educators and researchers the opportunity to investigate the amount of learning that can be attributed to each of the primary grades, K-3. The ability to link student Grade 3 standardized assessment outcomes to primary grade teachers could provide insight into the possible cumulative impact that teachers have on student achievement. Although the statistical assumptions are robust, this capability provides an empirical data point that can be used in evaluating teacher quality and effectiveness of instruction (Kane & Staiger, 2008; Rothstein, 2010).

All states employ some form of standardized tool to assess the academic achievement of students within their jurisdictions. In Indiana, the ISTEP+ is the tool used to measure student achievement in the subject areas of English/language arts (ELA), mathematics, science (Grades 4 and 6), and social studies (Grades 5 and 7). ISTEP+ describes student achievement levels according to the Indiana Academic Standards that were adopted in November 2000 by the Indiana State Board of Education. An applied skills assessment and a multiple-choice assessment, which are required components of the ISTEP+ program, are used to measure these standards once per year. Presently, test items are based on the Indiana Academic Standards identified by subject matter and grade level; however, the intent is to move toward assessing Common Core State Standards by 2014. This high-stakes test administered to students in Grades 3 through 8 continues to be tied to resources, accreditation, and school accountability (Indiana Department of Education, 2012b).

With the proliferation of technology and the speed with which we now access data sources, the ISTEP+ snapshot of student learning is administered in the spring and the results returned prior to the conclusion of the school year. The intent of the springtime administration is to provide student performance data that are reflective of what the student had learned during the school year. Algorithmic formulas calculating students' academic growth in comparison to their peers are now part of the statistical picture. Districts, teachers, and parents are certainly provided a broader set of metrics to consider in terms of student academic achievement. Tracking student gains after Grade 3 is determined with the students' academic growth score as well as achievement cut score, but unlike subsequent grade teachers, presently the Grade 3 teacher stands alone as being responsible for learning up to and through the third grade. A tacit assumption often comes with those Grade 3 scores that the student learning originates from and can be attributed solely to the Grade 3 teacher.

Statement of the Problem

The present student achievement and school accountability model, which utilizes ISTEP+ scores, does not take into account students' previous academic instruction and experiences (Kane & Staiger, 2008). Teachers are assigned students who possess attributes that vary in ways that may not be able to be measured by a standardized test score. Student factors such as intrinsic motivation, home environment, unmeasured prior experiences, and parent engagement most certainly impact student achievement gains. The configuration of a teacher's classroom may also reflect the principal's bias or preferential treatment of a colleague, ability grouping, and parent requests. The difficulty of measuring and then assigning levels of learning to specific teachers while controlling for these factors is problematic (Rothstein, 2010).

Significance of the Study

The significance of this study was to increase the research base that focuses on the link between teachers and student achievement. The findings have the potential to avail the principal, district leadership, and policymakers insight into the powerful impact routine decisions might have on learning in the building. For example, knowing the teachers in the school who produce the best readers would certainly shape the principal's teacher placement strategies. An additional benefit would be to direct and provide differentiated professional development that builds on teacher strengths or addresses weak instructional practice.

Having the knowledge that the second-grade teacher has more impact on student success in middle school than does the third-grade teacher would likely influence the decisions made by building principals and district-level administrators in how teachers are assigned to buildings and grade-level configurations. Policymakers would benefit from empirical evidence to assist them in investing the finite resources available to the district in the most effective manner. As a result of legislation throughout the country that connects teacher evaluation to student performance, providing a measurable connection between teachers in the non-assessed grades to student achievement in the assessed grades would provide further information as states or districts develop new teacher evaluation systems. Exploring the impact that first-, second-, and third-grade teachers had on Grade 3 ISTEP+ scores would add further insight into the national accountability movement and increase the body of relevant research regarding the teacher's influence on student learning.

Purpose of the Study

The purpose of the study was to explore and quantify the contributions to student achievement from first-, second-, and third-grade teachers utilizing Grade 3 ISTEP+ scores by analyzing a narrow cohort of students based on the following criteria: overall mean achievement scores in both math and ELA among the study sample; as school and first-grade, second-grade, and third-grade-level subgroups; and teacher subgroups within the grade level. The statistical results could be used as a non-experimental method for estimating teacher effect on student test scores.

Research Questions

The following questions were developed to guide this study.

1. Are there recognizable patterns of success in first-, second-, and third-grade teacher combinations in third-grade ISTEP+ scores? ISTEP+ scale scores were organized and summarized by using a descriptive model which included the following: the

mean, standard deviation, and standard error for combined math and ELA of the cohort; math and ELA separately for the cohort as a whole; and the mean, standard deviation, and standard error for each school, grade level, and classroom separately.

2. Are there significant differences in classroom mean ISTEP+ test scores among teachers at an individual grade level? This question was addressed with an analysis of variance model or ANOVA that used the *F*-ratio to test for classroom difference for each teacher at each school at each grade level. For example, was the mean difference in the third-grade ISTEP+ score significant if a student had been in Teacher A's first-grade class as opposed to Teacher B's first-grade class?
3. Are there patterns of success in first-, second-, and third-grade teacher combinations reflected in third-grade ISTEP+ scores? Do first-grade, second-grade, and third-grade teacher combinations predict a significant proportion of the variance in third-grade ISTEP+ scores? The question was addressed by using a simultaneous multiple regression models. A system of dichotomous coded variables identified the unique teacher combinations. The model used Y if a student was linked to a particular teacher or combination of teachers for the purpose of instruction and N if the student is not.
4. Are third-grade ISTEP+ scores predictable based on first-, second-, or third-grade teacher? This question was addressed using a simultaneous multiple regression model. Dichotomous coded variables Y or N identified a student's exposure to a particular teacher. An equation that provided the line that best described the relationship for the set of X and Y data were generated. Using the defined equation, a

relationship between teacher exposure and student ISTEP+ score was determined then analyzed.

Null Hypotheses

$H_0$1. There are no recognizable patterns of success on grade ISTEP+ scores that can be attributed to any single or combination of students' first-, second-, and third-grade teachers.

$H_0$2. There are no significant differences in third-grade classroom mean ISTEP+ scale scores among students assigned to specific teachers at any previous individual grade level.

$H_0$3. There are no patterns of success in first-, second-, and third-grade teacher combinations reflected in third-grade ISTEP+ scores. A significant proportion of the variance in third-grade ISTEP+ scores is not predictable using first-grade, second-grade, and third-grade teacher combinations.

$H_0$4. Third-grade ISTEP+ scores cannot be predicted based on first-, second-, or third-grade teacher.

Definitions of Terms

Beginning of the year adjustment is the redistribution of students in Grades 1, 2, and 3 to balance class size and/or comply with the teacher collective bargaining agreement (Evansville Vanderburgh School Corporation and Evansville Teacher Association, 2010).

Highly qualified teachers (HQT) are deemed highly qualified if the following criteria are met: (a) hold a bachelor's degree, (b) attain full state certification or licensure, and (c) show proficiency in each subject taught (U.S. Department of Education, 2012a).

ISTEP+ is the measure of student achievement in Grade 3 math and English/language arts (Indiana Department of Education, 2012c).

Official enrollment or pupil membership date is the date that an official count of students in a school is determined and reported to the Indiana Department of Education (Indiana Department of Education, 2012b).

Random distribution is a randomized placement of students in classrooms based on student–teacher ratio and gender balance (Rothstein, 2010).

Socioeconomic status is "the economic and sociological combined total measure of a person's work experience and of an individual's or family's economic and social position in relation to others, based on income, education, and occupation" ("Socioeconomic status, 2012, para. 1).

Teacher, for the purpose of this study, is defined as a certified professional employee who provides instruction to students.

Title I status is the economic circumstance that identifies students at or below the poverty level as defined by the federal government (U.S. Department of Education, 2012a).

Assumptions

The assumptions of this study existed in the following manner:

1. Students were randomly assigned to teachers.
2. The ISTEP+ was administered to students with fidelity and based on standards for ethical practice.
3. The ISTEP+ state assessment questions did not change during the years of the study.
4. The average age of the students whose data were used was nine years old.
5. Data used in the study were selected from official EVSC reports submitted to the Indiana Department of Education.
6. Data were linked together using a common date.

7. Students received a standard curriculum across all schools and at each grade level (Indiana Academic Standards).
8. Teachers were *highly qualified* to teach each grade level set of standards.
9. Teachers in each school and in each grade level were exposed to the same professional development in regards to curriculum alignment, data analysis, and effective instructional practice.

Delimitations

Delimitations of the study existed in the following manner:

1. The existing data available from the Evansville Vanderburgh School Corporation (EVSC) during the time period 2008 through 2011 were used for this study.
2. Elementary schools with the grade level configuration of kindergarten through Grade 5 from the EVSC were included in this study.
3. The principal of each building remained the administrator throughout the length of the study.

Limitations

Limitations of the study existed in the following manner:

1. Controlling for a precise teacher-student link, the population of the study included EVSC first-grade students during the year 2008-2009 from a possible pool of 20 elementary schools. Diversity of schools identified for the study reflected population demographics present in the district at large.
2. Students were randomly distributed among teachers in each grade level within the building. EVSC district-level administrative practice provided building level principals the autonomy to assign student to particular teachers.

3. School socioeconomic status, culture, climate, policy, and procedures may vary among the schools.

Summary and Organization of Study

It is clear and supported by considerable research that teachers have a profound impact on student achievement. Learned educators as far back as John Dewey have recognized that connecting with students to stimulate and be a catalyst for authentic learning is a multifaceted human collaboration (Deci & Ryan, 1985; Dweck, 2000; Marzano, 2003; Wiggins, 2010). Within the learning paradigm, all participants bring with them a confluence of cognition, emotion, past experiences, and innate tendencies. Even though the teaching and learning interaction is complex, it can be assessed and the contribution of each teacher measured. Our society requires that children be provided with a foundation of skills, knowledge, and mental processes that will ensure their future success and it is our responsibility to provide quality teachers to do just that. There are substantial implications in terms of education policy, political power, and resource distribution. The review of the literature establishes the widely used link between standardized test scores to school attainment, earnings, and aggregate economic outcomes (Angus & Mirel, 1999; Hanushek, Jamison, Jamison, & Woessmann, 2008).

This study is divided into five chapters. Chapter 1 provided an introduction for the study, a statement of the problem, the purposes of the study, research questions, null hypotheses, definition of terms, assumptions, delimitations, limitations, and a summary and organization of the study. Chapter 2 provides a review of the literature that summarizes the evolution of schooling from 20th century educational reform movements to current efforts to link teacher accountability to student academic growth and achievement. Chapter 3 describes the methodology and design of the study; population parameters, source of data, and methods of analysis. Chapter 4 states the findings to questions posed in Chapter 1. Chapter 5 presents a summary of the findings, conclusions, and a discussion of immediate implications and suggestions for further study.

CHAPTER 2

REVIEW OF RELATED LITERATURE

The goal of this study was to provide additional information to add to the current body of research intended to inform educators, policymakers, and the general public when making decisions that impact public education. The review of the literature is organized in five parts: an introductory rationale for the public's increased interest in student academic achievement; a summary of the foundation of our present public education model; a discussion of significant reform movements throughout the 20th century; and an examination of topics that currently surround educational debate, including teacher impact on student achievement, teacher quality in terms of performance measures, innovation and its funding both public and private, and accountability for student achievement.

Searching for the silver bullet to "fix" public education is a multi-billion dollar enterprise. Big money intimately involved in a variety of school reform enterprises and the rationale behind the $5 billion taxpayer financed *Race to the Top* has caused a tidal wave of innovative initiatives and unrelenting national discussion (U.S. Department of Education, 2009). Even though there are rules of compliance, district leaders welcome the potential additional funding. The dollars being pumped into promising models come from philanthropists like Bill and Melinda Gates and charitable groups like the Wallace Foundation and Lilly Endowment. The Fordham, Brookings, and American Enterprise Institutes are also integral voices in the educational policy discussion and private reform movements. Integral voices, such as Ladd, Noguera, and Payzant's (2011) A Broader Bolder Approach to Education, Michelle Rhee's (2010) Students First Organization, and Jeb Bush's (2010) Foundation for Excellence in Education, have gained traction attracting scores of members who prescribe to a specific educational agenda. These entities are committed to keeping the pressure on those who seemingly support the status quo. A shift toward econometric research has also become a guiding force in the business of school reform. Noted education commentators John Chubb, Terry Moe, Eric Hanushek, Fredrich Hess, and Tony Wagner are examples of the influential voices that tie economic results to student achievement outcomes. The proliferation of commentary based on research and philosophical predisposition has found an audience among citizens across the intellectual and social spectrums.

The dominant emphasis continues to be characterized by advocating structural changes for systemic reform, including curriculum reconstruction, Common Core Content Standards (National Governors Association, 2009), and performance pay, none of which currently have been effective in penetrating the operational and isolationist siloed school systems (as cited in Chubb & Moe, 1990; Hanushek, 1992; Hattie, 2003). Some of the reform efforts have taken hold and several have shown relative increases in student achievement measures reflected in high-stakes tests. Many of the reforms, however, have not brought about the anticipated increases in student learning required to equip children for the rigors of life after high school. The intense scrutiny of the practices and policies associated with education both public and private is a worldwide phenomenon driven most notably in a climate of outcome-based economic rationalism. Because political, economic, and industrial issues surrounding educational effectiveness are sensitive and complex, the relevance of teacher quality and quality teaching intended to equip students adequately to meet the constantly changing demands of modern society and the workplace have never been more important or widely debated (Ingvarson & Rowe, 2007; Schlechty, 2000; Wagner, 2008).

In order to enter the debate, it is essential that learning and achievement be measured in a standardized and understandable fashion. Parents and politicians make the assumption that the standardized tests administered in schools adequately measure the right things: content, skills, and processes that will enable children to become educated and productive citizens (Guskey, 2003; Marzano, 2005; Reeves, 2006; Schmoker, 2006, 2011). Because academic accountability is everywhere, from neighborhoods of poverty to those of affluence, a sense of urgency has been created (Data Quality Campaign, 2012). State legislators and school districts within the states are interested in solutions. As a result, educational entrepreneurs are encouraged to innovate and provide options that may enhance student achievement. In 2009, the consortiums, such as New Profit, Public Impact, the Center for American Progress, and the American Enterprise Institute who are often on opposite sides of the philosophical vision for improving education, collaborated on *Stimulating Excellence* (Center for American Progress, 2009). The report advocated for a more robust support from states and federal agencies to provide funding and flexibility for innovation. It also identified very prescriptive models implemented across the country that held promise of higher student achievement. Touting a research foundation supported by empirical evidence, the System for Teacher and Student Advancement (TAP; National Institute for Excellence in Teaching, 1999) and Knowledge is Power Program (KIPP; Henig, 2008) are two programs which have gained nationwide attention. Each of these programs placed emphasis on accountability and teacher quality as foundations for their success. Both highly prescriptive programs include significant professional development in lesson design, classroom management, and engagement.

Adopting selected ideas from research has been helpful, but bringing the model to scale and applying it to all schools in a community regardless of context has failed to bring about the increased learning across the board. For example and as a direct response of the 2002 No Child Left Behind (NCLB) legislation, the Reading First model was highly marketed as the method that would provide all children the foundational skills essential for reading (U.S. Department of Education, 2001). This program focused on implementing scientifically based strategies intended to ensure effective early reading instruction in classrooms. Through Reading First, states and districts received funding and technical support to apply the instructional and assessment tools consistent with reading research. The intended result was that all children would learn to read well by the end of third grade (U.S. Department of Education, 2012b). The Reading First model was identified as the required model for many schools that were considered failing under the NCLB accountability policies developed by the Department of Education. As part of the model, the five big ideas of reading became an integral component of early reading instruction in primarily Title I schools: phonological awareness, alphabetic principle, fluency, comprehension, and vocabulary (National Reading Panel, 2000).

The remainder of the Chapter 2 examines the notion that student achievement depends upon many factors, both within the confines of the school building and outside. Because learning is a personal experience, creating the optimal conditions that lead to academic success require a more indepth examination of its process. It is becoming increasingly more apparent that in order to provide those conditions for each child, education as we know it much change. One size does not fit all (Weisberg, Sexton, Mulhern, & Keeling, 2009).

Foundations of our Present System

For thousands of years, it has been evident that those who hold the knowledge and information have the power (Bacon, 1605). In emerging simple social structures, only a select few were fortunate enough to be the owners and purveyors of knowledge. Kings, religious leaders, and their inner circle of associates enjoyed the benefits of being educated and informed. They brandished that associated power by lording over the ignorant masses. As communities began to trade with others, commerce evolved and so did the opportunities for shared information and knowledge. Universities and libraries were built and those affluent enough to attend continued to hold the power. Even in the early years of our nation's expansion, the more prosperous citizens had access to information because education continued to be an expensive commodity (Robinson, 2010).

The roots of America's public education can be traced back to the Puritan and Congregationalist schools in the 1600s. Religion was a catalyst in identifying a purpose for more educated communities. Teaching children to read the Bible was the church's ultimate goal (Grelbrick, 1999). Common school reformers saw great value in the concept of educating as many people as possible and their commitment provided a framework for the availability of free elementary education in the 1800s. The instructional delivery model reflected the organizational systems applied by emerging industry at that time. Children as the *raw* material matriculated through a system in which teachers provided the knowledge and students absorbed it. The finished product, if a high school graduate, was then certified as educated.

Education was regional and communities basically agreed upon the skills, processes, and knowledge that students needed to know in order to be a productive citizen in the local context.

The student's job was to be compliant, accept information, and retain the knowledge provided by the classroom teacher. Assembly-line education was the norm (Trilling & Fadel, 2009).

This factory-like structure flourished in the early 1900s. It was important for the general welfare of the country that large masses of people obtained a basic set of skills and knowledge. Providing students with an experience requiring a high level of cognitive demand was not the purpose of schools at the turn of the century. If students were not successful in the classroom configuration of the time, they would fall behind and then most likely drop out. The economy did not require a large number of highly educated individuals because semi-skilled, non-skilled jobs were abundant and farm work continued to be available for those who progressed through the system or dropped out (Trilling & Fadel, 2009).

Looking at the larger enterprise of education, there was little or no connection between the practice of teaching and the result of learning. The prevalent consensus was that some students were capable of learning and some simply were not. Tracking students into vocational studies was a common and appropriate practice. Many parents and teachers remained confident that a useful education was one that prepared a child to earn a living, vote intelligently, spend wisely, and be a good citizen (Hampel, 2008). Even though that same definition of education is relevant today, there is a profound difference in the world in which children now live.

Today's scholars, intellectuals, and theorists consistently point to antiquated content and processes that constitute current public education (Hanushek et al., 2008; Hess, 2009; Robinson, 2010; Sprenger, 2010). Students can create their realities and retrieve facts instantly through technology. Teachers are no longer the keepers of the knowledge. The definition of a useful education may be consistent with the past, but in the opinion of most analysts, the content expectations of schooling have changed significantly. As Pink (2005) suggested different kinds of thinkers were successful in previous decades.

> A certain kind of person with a certain kind of mind—computer programmers who could crank code, lawyers who could craft contracts, MBAs who could crunch numbers. The future belongs to a very different kind of person with a very different kind of mind—creators and empathizers, pattern recognizers, and meaning makers. (Pink, 2005, p. 1)

School Reform Movements

The evolution of public education in our country has experienced three significant school reform movements since 1900. With the understanding that information and knowledge were necessary underpinnings for a productive viable nation, the first school reform movement began with the spread of compulsory school attendance. In the United States, the first state to require compulsory attendance was Massachusetts in 1852 when Horace Mann helped establish a statewide system of education. The last state was Louisiana in 1917 (National Conference of State Legislatures, 2012). Social reformers saw this as an opportunity to educate more students at higher levels. During the decades prior to World War II, reformers advocated and succeeded in increasing the number of school attendance days from 144 to 174. Student absenteeism decreased and high school graduation grew from about 6% to 29% (U.S. Bureau of the Census, 2011). In the years following the World War II, college degrees granted by institutions of higher education improved from 27,410 to 122,484 (U.S. Bureau of the Census, 2011).

The Second Reform Movement

After World War II, a second educational reform movement emerged as service men and women were given the opportunity to attend college or vocational school through the

Servicemen's Readjustment Act of 1944, more commonly known as the GI Bill (U.S. Department of Veterans Affairs, 2012). This omnibus bill provided schooling for the parents of the present baby-boomer generation. The promise of a good life included a solid education, and the United States was committed to public education as the means of delivering on that promise.

After the war and into the 1960s, schooling was largely subject-matter centered and the federal government infused a considerable amount of money into the development of curriculum, teacher training, and research. Opportunities were available and the majority of Americans were in the position to take advantage. This system of educational opportunity directly supported the foundation of America’s middle class where college enrollment became more than a wish; it developed into an expectation (Hampel, 2008).

Urgency Drives Reform

The Cold War and the successful launching of the Soviet satellite Sputnik in October of 1957 created an increased sense of urgency regarding the necessity of sustaining an educated populace. An uneasiness regarding America’s place as the democratic superpower appeared to be threatened. The discussion involving education changed from a societal issue to one of a matter of national security.

Through focus and funding, the federal government inserted itself more boldly into what had historically been a local or state function, public education. To meet the country’s need for a more broadly educated population, different approaches were required. *The Pursuit of Excellence,* published by the Rockefeller Brothers Fund in 1958 (Hurd, 1958), was just one of several important reports published during the Cold War that emphasized more rigorous content expectations. There was a shared belief that the United States was not producing sufficient scientific minds. A series of reports that focused primarily on science education indicated that the technological breakthroughs were happening at an unprecedented rate (Hurd, 1958). This phenomenon produced significant challenges for public education and that fact necessitated a change in current educational practice. There was general agreement that the nation required an "adequate supply of highly trained scientists, mathematicians, engineers, the society also needed a highly educated citizenry that understood the scientific enterprise" (De Boer, 2000, p. 586). The response to Sputnik precipitated an intentional effort to enhance the disciplines of science and math in classrooms across the nation. Some educators of national prominence (Carnegie Commission, 1991) cautioned that as schools turned out technically trained individuals, they needed to create a balance that also supported science literacy as an intellectual achievement. Discovery and exploration, according to Hurd (1958), was an essential component of education.

In the realm of classroom instruction, two distinct hypotheses began to emerge. One was predicated in Bruner's (1960) book, *The Process of Education,* whose approach supported a learning process method or learning-how-to-learn emphasis rather than a curriculum designed in subject matter silos. Many teacher education programs during the 1970s endeavored to train teachers in the discovery or inquiry method of instruction that was based on Bruner's design and implications made by Hurd (1958). The curriculum and instructional design integrated the content disciplines so that students re-encountered concepts in more complex forms as they moved through the grade levels. The contrasting approaches, however, revealed separate and isolated content disciplines and continued to be the most prevalent model of instruction practiced in public schools throughout the nation.

Typical 1970s teacher education programs reflected Bruner's and many of John Dewey's progressive education concepts (Mayhew & Edwards, 1936). Preservice teachers learned about the importance of the emotional, artistic, and creative aspects of human development (Westbrook, 1991) and how those concepts applied to teaching and learning in the classroom.

It became clear, however, that the veteran teachers in the classrooms were reluctant to embrace Bruner's (1960) model. Typically student teachers met with supervising teachers who respected seat time, student compliance, memorization, and recall instructional delivery methods. Assembly line instruction remained thoroughly entrenched in public school classrooms. In most cases the integrated model, if attempted, did not replace traditional classroom practice. In typical classrooms across the nation, the majority of teachers did not embrace the instructional model that most likely created the environment for integrated learning (Reeves, 2006). Whether the new design held promise or not, classroom instruction continued to proliferate the status quo of uneven learning opportunities from child to child or school to school (Wong, 2004). The brighter, college-bound students were afforded more intellectually demanding experiences; the students who were going into the workforce directly from high school were prepared at a less rigorous level.

The topic of student achievement and quality schools continued to dominate public interest and policy during this time. In 1979 the Department of Education became a cabinet-level department of the U.S. government under the Carter administration and began operating in May of 1980 (Department of Education Organization Act, 1979).

The Case for Accountability

In 1983, a prominent report on American education ushered in the third major reform movement (Bell, 1983). The report publicized a number of alarming statistics that refocused attention on public education. Bell's (1983) report successfully stressed the need for higher levels of learning by all students. This resulted in the mobilization of a wide range of societal interest, resources, and expertise to address educational issues (Wong, 2004). The popular notion was that a high school graduation rate of 50% could not be tolerated if our country was to survive. Just as Sputnik created the awareness among the public that education could be a matter of national security, so did *A Nation at Risk.* The report required that changes be made in the educational system and proposed that educators be held accountable for student performance (Wong, 2004). The job market demanded increased knowledge and more sophisticated levels of performance from public school students (Bell, 1983).

The report commissioned under President Ronald Reagan in response to the performance comparisons between U.S. students and students in other industrialized countries did not contain an accountabilty mechanism to measure performance (Bell, 1983). The report was successful in again bringing urgency and significant national attention to public education. Not since October 1957 had so much attention been focused on schools and the quality of curriculum and instruction (Angus & Mirel, 1999). This position was advocated by the intellectual as well as the blue collar communities. The reform movement was termed *standards based* by the former and *back to the basics* by the latter constituencies. Regardless of the terms, both identified a common purpose that emphasized basic subjects such as reading, writing, and arithmetic along with a strong work ethic.

Standards for content and performance. The first federal reauthorization after A Nation at Risk was the Hawkins-Stafford Amendments of 1988. The amendments emphasized standards-based reform that created momentum for subsequent legislation. The Improving America's Act of 1994 changed the nature of accountability in federal legislation based solely on rules and regulations that focused on fiscal procedures such as budgeting and uniform accounting systems to accountability based on student performance (Wong, 2004). States were encouraged to develop more comprehensive systems that would define and measure student achievement in terms of adequate yearly performance (AYP).

In 1989, President George H. W. Bush hosted an education summit in partnership with the National Governors' Association. The intent of the gathering was to set broad national performance goals to be achieved by 2000 (Wong, 2004). Following the summit, a National Education Goals Panel was established to identify academic goals, student assessments, and standards of performance (Kendall, 1996). The goals panel identified standardization of content and achievement as the primary objective. The panel also emphasized diversity and the necessity of moving curriculum and instruction into the age of technology.

Identifying the knowledge, skills, and processes that would be needed and then retooling the educational system to deliver was fraught with conflicting and contradictory ideas and pathways to meet the objectives. The meaning of rigorous content and process standards was impossible to reconcile among drastically divergent educational philosophies. The search for effective policy that honored the range of community needs became increasingly more complex. There was a broad consensus in the education community, however, that preparing students for the challenges they would face in the future could not be achieved by applying solutions based on traditional classroom practice. Ironically, the actions implemented by many communities and states tightened the curriculum and began testing students on discrete and narrowly focused content more frequently (Berliner, 2010).

Research meets pedagogy. Concurrent to the larger educational reform movements, social scientists such as Rita and Ken Dunn, Michael Posner, and Howard Gardner were designing and carrying out research certain to influence the teaching and learning debate (Sousa, 2010). Long-held assumptions about who could learn, how much could be learned, and what could be done if students did not learn emerged as topics of interest for a host of stakeholders: legislators, educators, and the general public. The research on teaching surfaced as a topic of particular interest. Even though John Dewey's Laboratory School at the University of Chicago had been in place since the early 1900s, for the most part, education had not been the beneficiary of research-practice partnerships (Hinton & Fischer, 2008).

Persuasive research. One of the most influential reports in relation to teaching and learning was submitted in 1966 as a response to the Civil Rights Act of 1964. The U.S. Department of Health, Education, and Welfare commissioned a study to determine if schools and/or schooling conditions had a significant bearing on student academic achievement (Coleman et al., 1966). Public schools were relatively uniform in their configuration and structure throughout the country; one teacher was responsible for a classroom of children. Compulsory school attendance began in Grade 1 and continued through Grade 12 or until a student reached 16 years of age. The Coleman et al. (1966) study was intended to determine if equity existed in schools or if there were other more significant factors that influenced levels of learning and student academic achievement. As a result of the study, the socioeconomic status (SES) or circumstances of birth was identified as the most significant predictor of academic achievement. The impact of the interaction within the classroom between the teacher and his or her students was not cited as a major predictor of student success. Since student−teacher relationships and classroom instruction had not been recognized as significant, teachers were often evaluated on whether they were pleasant, well dressed, knew about their subjects, and were their classrooms quiet and orderly (Danielson & McGreal, 2000). The predominant and overarching goal in the classroom was to create upstanding citizens. Tacitly, academic achievement was not expected from every child, especially from those coming from impoverished backgrounds.

Early findings. The principal implication that surfaced from the Coleman et al. (1966) study was that schools as they were typically designed had very little impact on student achievement. Student demographics were the deciding factor that influenced learning (Coleman et al., 1966). As part of *The Great Society* initiatives promulgated by the Johnson administration, federal dollars poured into programs that were created to attack the root cause of non-achievement; primarily factors associated with poverty and race (Wong, 2004). Even today through the Title system, federal dollars continue to be provided to districts in large part based on the same criteria established during the Johnson administration. And with any type of governmental funding, rigorous sets of standard conditions of compliance accompany the dollars with little or no consideration for the unique needs of the community the funding is intended to support.

The general interpretation gleaned from the Coleman et al. (1966) report contended that the influence of school environment was not considered a significant contributor in producing increased achievement. This interpretation woven into the fabric of educational debate and analysis among policy makers and educators impacted the potential for higher levels of learning significantly (Rivkin, Hanushek, & Kain, 2005). The unintended consequences associated with this study were striking: if a child had the capacity to learn supplemented by an external supportive environment then he or she would achieve academically. Among teachers, parents, and students themselves, the predetermined outcomes for achievement had been set before students even walked into the classroom. If the child did not learn, it was attributed to factors other than school culture, misaligned curriculum, or ineffective instruction (Lakoff & Johnson, 1980). The consensus was that factors outside of the school itself inhibited student learning (James, 2007). Jencks et al. (1972) also corroborated Coleman et al.'s (1996) findings with the assertions that typically schools did little to lessen the gap between rich and poor students or between students who appeared to be more academically able and those who appeared to be less able. The report offered little evidence that school reform efforts improved academic achievement between students who lived in environments that offered opportunities to build background knowledge and those who did not have those advantages.

Additional studies and commentary by noted educational researchers confirmed the general interpretation of Coleman et al.'s (1996) research correlating student learning to student background circumstances: "The higher the SES of the student's family, the higher the academic achievement" (Boocock, 1972, p. 32). "The positive association between school completion, family socioeconomic status, and measured ability is well known" (Welch, 1974, p. 32). "SES predicts grades, achievement and intelligence test scores, retentions at grade level, course failures, truancy, suspensions form school, high school dropouts, plans for college attendance, and total amount of formal schooling" (Charters, 1963, pp. 739-740).

So strong was the acceptance of the relationship between academics and student home circumstances, the claim was consistently cited and rarely questioned by researchers and practitioners. In fact, the research findings varied considerably (White, 1982). White (1982) conducted a meta-analysis of 200 studies that provided correlations between SES and student achievement. The scope of the study revealed over 600 correlations, identifying four elements associated with SES that related to a high degree with student academic achievement: income, home atmosphere, and parent education and occupation all showed effect sizes of .38 or above. Home atmosphere presented the strongest link to student achievement with an effect size of 1.42

(White, 1982). White's study did support the prevailing belief that achievement was largely associated to SES, but he also asserted that home environment explained more variance in achievement than did the other three factors combined. Fan and Chen (2001) also found that the home environment produced a .33 correlation on average and 10.89% of the variance in student scores in their meta-analysis of parent involvement and their association to student achievement.

The meta-analysis by Hattie (2008) and additional individual studies by White (1982) and Fan and Chen (2001) certainly were not the only investigations into the student achievement issue. Intellectuals and policy analysts associated with the National Bureau of Economic Research, the National Board of Professional Teaching Standards, and Urban Education Partnerships along with philanthropic foundations embraced the need for further investigation. Adding to the body of research, studies began to emerge especially from the social sciences linking teacher and student expectations of learning with a student's level of academic achievement (Kamins & Dweck, 1999; Sanders & Rivers, 1996). As stated earlier, a significant number of teachers carried with them the belief that they had little bearing on individual student learning. Classroom instruction really did not matter one way or the other because the student's capacity to learn was set by factors independent of teacher-student interactions (Bornholt, 1991).

Lasting implications. For the next 20 years little changed in terms of increasing the percentage of students who achieved at higher levels. The only break in this pattern was accomplished through efficacious teachers, such as Jamie Escalante, the charismatic former East Los Angeles high school teacher who taught the nation that inner-city students could master calculus (Barth, 1999). Stories like this and incidental academic success in the midst of poverty were celebrated but deemed anomalies. A list of *beat the odds* schools compiled by the Education Trust in 1999 (Barth, 1999) surveyed to determine commonalities and characteristics among those schools. Along with aligned curriculum and home atmosphere, the importance of the teacher–student interaction proved to be significant (Barth, 1999). Unless struggling students were fortunate enough to be with teachers who applied effective pedagogy and believed they truly impacted student achievement, the Coleman effect continued undiminished (Marzano, 2008; Rosenthal & Jacobson, 1968). The predominant belief continued to be that teachers were not of significant importance in increasing student achievement (Boocock, 1972; Bornholt, 1991; Charters, 1963; Coleman et al., 1966; James, 2007; Jencks et al., 1972; Lakoff & Johnson, 1980; Welch, 1974).

Although John Dewey's work in the early 1900s had provided ample insight into the probable connection between the quality of instructional design and delivery as influential in student achievement, the true impact of teachers on student learning had yet to be definitively acknowledged. Based upon John Dewey's attention to educational theory and practice, much of the subsequent research regarding interaction in the classroom could be traced back to his work.

As researchers began to look more closely at the contributions of teachers in relationship to student acquisition of skills, the national narrative was adding important content to the body of research findings. Subsequent studies were linking student academic success with specific identifiable teacher *behaviors*. A move toward a deeper analysis of the teacher-level impact (Marzano, 2003) continued during the 1980s. For the purpose of analysis, teacher behaviors were defined as decisions made by classroom teachers regarding instruction. Much of the educational research during the 1990s focused on identifying specific teacher behaviors that could be termed as effective or ineffective in relationship to student achievement. Because assessing the impact of isolated behaviors was neither prudent nor to this point very successful, researchers designed their studies to group behaviors into common categories. Instructional strategies, classroom management, curriculum design, and assessment tended to be the common groupings for study (Cotton, 1995; Creemers, 1994; Danielson & McGreal, 2000; Marzano, 2000).

A significant study by Wright et al. (1997) measured the difference in student achievement based on effective versus ineffective teachers. The following year an additional study by Sanders & Rivers (1996) measured the cumulative effect that teachers had on students. The study concluded that one year with an ineffective teacher was not debilitating, but the cumulative effect of several ineffective teachers in a row could be disastrous (Sanders & Rivers, 1996). There was now a common belief that the connection between teacher actions and student achievement had been made (Darling-Hammond, 1997; Goe & Holdheide, 2011; Good, 2010; Hattie, 2003; Marzano, 2005; Nye et al., 2004; Sanders & Rivers, 1996).

Research and a Shift in Teacher Efficacy

Historically many have noted the concept of self-determination as a significant catalyst an individual's motivation to achieve. Gandhi (n.d.) observed, "If I have the belief that I can do it, I shall surely acquire the capacity to do it even if I may not have it at the beginning." Bandura's (1986) theory of self-efficacy explained and built on the notion of self-determination. His theory of *reciprocal determinism*—a person's behavior both influences and is influenced by personal factors and the social environment—was an important component in behaviors linked to achievement. The theory correlated with teachers' classroom attitudes and interactions with students. The back-and-forth feedback between student and teacher directly impacted the student's belief in his or her ability to achieve. Studies supported the theory that if teachers perceived they had the capacity to increase student achievement, they exhibited behaviors that were more likely to produce that outcome (Dweck, 2000; Good, 2010; Kamins & Dweck, 1999; Rosenthal & Jacobson, 1968).

Recent research in student motivation supported the concept of efficacy as a contributor to increased achievement as well as the lack of achievement (Blackwell, 2007; Dweck, 2000, 2006). It has been posited that if students and teachers believe that intelligence is fixed and that each person has a finite amount or that factors outside their control dictate achievement, then the confidence that either student or teacher can affect higher levels of learning is diminished (Wiggins, 2010). This fixed mindset has been accepted especially when applied to students who live in poverty. President George W. Bush characterized the fixed mindset as “the soft bigotry of low expectations” in a Florida speech in 2004 after the passage of the NCLB act (as cited in Noe, 2004, para. 1).

Many social and neuroscientists agree that expectations and mindsets impact both student and teacher behaviors (Marzano, 2010; Willis, 2010). In one study with preschoolers, Cimpian (2010) measured the impact that subtle praise had on student motivation to persist. The result of the study indicated that there was no impact prior to a student making a mistake. Once the student encountered difficulty, however, the teacher’s comments impacted student motivation to persist in a positive or negative way. Students who were told initially that they were “good drawers” were much more likely to feel sad and fail to look for strategies to fix the mistake than students who were told that they “did a good job of drawing” (Cimpian, 2010, p. 315). Cimpian (2010) additionally investigated children’s motivation in relationship to generic feedback.

Cimpian (2010) further studied the student’s exposure to generic language about his or her ability to perform. The results of the study indicated that generic language feedback was detrimental to children’s motivation when compared with exposure to analogous non-generic language.

> The fact that a manipulation involving only a few generic versus non-generic sentences (whose content was otherwise identical) had such a dramatic impact on children's behavior and attitudes toward a novel activity is a compelling illustration of the power of this linguistic distinction. (Cimpian, 2010, p. 1337)

Cimpian's investigations of verbal feedback to students supported research performed by Dweck (2006) regarding the influence a teacher's mindset has on his or her behavior and subsequently on student efficacy. This belief had the ability to undermine both high achieving as well as struggling students. Students who were identified as having an above average IQ worried that failures were unacceptable because they were *bright* and should therefore be able to achieve without effort. Challenges or missteps would undermine their confidence about the amount of intelligence they possessed. Similarly, a less than average IQ proved to create a false ceiling for both students and teachers in terms of achievement levels (Mueller, 1998). In both cases, the fixed mindset gave credence to the notion that innate and external factors determined the student's level of learning. Wholly, Dweck argued that if a student believes that work and effort can improve his or her work or standing then he or she is much more likely to persist and therefore to succeed. If a student believes that his or her work is a reflection of his or her ability, then his or her confidence and motivation can be weakened (Dweck, 2000; Mueller, 1998).

With the help of educational neuroscience, it is now known that intelligence is not fixed but is something that can be cultivated through effort and education (Sousa, 2010). Educational neuroscience and neuroimagery are an evolving scientific field that merges researchers in cognitive neuroscience, developmental cognitive neuroscience, educational psychology, educational technology, education theory, and other related disciplines to explore the interactions between brain functions and education. Currently, researchers in educational neuroscience have investigated the neural mechanisms associated with many of the brain processes linked to reading, math, genders, demographics, and learning difficulties (Coch, 2010). Educational neuroscience intentionally connects research that has the potential to inform educational practice (Posner, 2010; Sousa, 2010).

The brain has plasticity (Willis, 2010), and teachers are in the profession of changing the human brain every day (Sousa, 2010; Wiggins, 2010). Neuroscience appears to have the ability to identify changes in the brain that occur through an individual's effort and practice. The study of the brain does strongly suggest that individuals can improve their abilities with focused hard work. In short, students and teachers with a growth mindset believe that intelligence is a potential that can be realized through learning (Sprenger, 2010). Costa and Kallick's (2008) work surrounding habits of the mind explained that implementing classroom strategies that employed a focus on work ethic and perseverance within a growth mindset had significant positive impact on student achievement. As a result, confronting challenges, profiting from mistakes, and persevering in the face of setbacks become ways of getting smarter (Costa & Kallick, 2008; Dweck, 2006; Kamins & Dweck, 1999).

Accountability and Funding

By the mid-1990s, state and national legislatures formally tied public K-12 education funding to the requirement that progressively more students would achieve minimum competency levels each year. This requirement was articulated in the Goals 2000 report commissioned through the Educate America Act (U.S. Department of Education, 2011). The link between accountability and achievement became stronger and legislatures had the power to reduce school funding if students did not perform up to expectations. The most potent lever for change, however, proved to be the practice of publicly identifying schools and districts that succeed or fail to make consistent AYP. By the end of the 21st century, states had begun to develop and adopt rigorous academic and accountability standards for both students and educators. This movement was given increased credibility in 2001 through the passage of the Elementary and Secondary Education Act, more commonly known as No Child Left Behind (Goe & Holdheide, 2011).

Drawing on pertinent research, it was determined that within schools, teachers were the single most influential factor associated with student achievement (Darling-Hammond, 1997; Goe & Holdheide, 2011; Hattie, 2008). Hattie's (2008) meta-analysis examined 58 studies that focused on factors related to teaching and learning. By using effect size, it was possible to use Hattie's analysis to these interventions and their probable impact on student learning. In a rank order list of effect sizes, teacher behaviors, instructional strategies, and classroom organization had a larger impact on student achievement than did student SES or home environment (Wiggins, 2011).

The overwhelming evidence provided by an increasing number of studies has linked increased student achievement with the classroom factors involving the teacher (Anderson, 1979; Darling-Hammond, 1997; Haycock, 1998; Marzano, 2010; Marzano, Pickering, & Pollock, 2001; Nye et al., 2004). This connection was made in the strongest terms by noted econometric researcher, Hanushek (1992) who noted, "The difference between a good and a bad teacher can be a full level of achievement in a single school year" (p. 3). Adding further support to Hanushek's conclusion, Harvard and Columbia researchers, Chetty et al. (2011), in a paper for the National Bureau of Economic Research, concluded that "a high value-added teacher makes a student more likely to go to college, have higher salaries, and live in higher SES neighborhoods" (p. 1). According to Weisberg et al. (2009), The New Teacher Project, a teacher's effectiveness, the most important factor for schools in improving student achievement is not measured, recorded, or used to inform decision-making in any meaningful way. The University of Tennessee Value-Added Research Center was created to oversee the development of an evaluation system that would efficiently and effectively determine an individual teacher's influence on the rate of academic growth for student populations (Sanders & Horn, 1998). The Tennessee Value-Added Assessment System (TVAAS) was designed to produce a measurement that extended above and below grade level, collected longitudinal data, and used multivariate statistical analysis that would produce unbiased and efficient estimates of teacher effectiveness that resulted in student achievement. Using TVAAS data that encompassed approximately 3 million records of Tennessee's entire population of Grades 2-8 as the measure, the study concluded that one of the highest factors correlating with different rates of growth among students was the instructor (Sanders & Rivers, 1996). Haycock (1998) wrote extensively on the importance of great teachers. Citing evidence from a series of studies, Haycock further supported the importance of well-prepared classroom teachers. Teachers' proficiencies in deep content knowledge and pedagogy combined with experience are all important elements that characterize great teaching and teachers (Darling-Hammond, 1997; Haycock, 1998).

Because there continues to be a significant level of uncertainty about the reliability and validity of how a teacher's impact is measured, many teachers still subscribe to the notion that their performance is affected by the students assigned to them. Based on the students they teach, teachers might enjoy an advantage or they may be at a disadvantage in terms of effecting student achievement (Rothstein, 2010). According to Bandura (1986), "The beliefs that adults hold are the best predictors of future decisions and behaviors. People regulate their level and distribution of effort in accordance with the effects they expect their actions to have" (p. 129). Drawing from Bandura's theory and Dweck's (2000, 2006) research as the foundation, the "Coleman Effect" had and continues to have an impact on students. With the increased focus on accountability for each child's learning, teachers who have fixed mindsets that are reflected in low academic expectations regarding entire groups of students are experiencing a heightened degree of anxiety about their job performance.

Economics Drives Accountability

Econometrics has emerged as a measure of linking economic impact with schooling. This research has become a force for educational reform in recent years. In searching the Internet using the key words *education* and *economics*, topics relating to the cost of poor education and the astounding economic impact it has had on the country are numerous. Econometric research is now one of the factors used to craft education policy. There are enormous societal and economic costs associated with education in general and high school graduation in particular (Hanushek et al., 2008). The Alliance for Excellent Education (2010) is only one organization that provides a factsheet that exposes the high school dropout concerns and their financial and social costs to communities, states, and the nation. Estimates suggest that 3 trillion dollars were lost in the U.S. economy as a result of lost income opportunities, chronic unemployment, incarceration, and the need for government assistance (U.S. Department of Education, 2010b). It also has been estimated that if dropouts from the class of 2009 had graduated, the nation's economy would benefit from nearly $335 billion in additional income over the course of their lifetimes (American Psychological Association, 2011**).** Additionally the

U.S. Department of Education under the Obama administration has directed considerable funding be available for turning around dropout factories (U.S. Department of Education, 2010b).

Assessing Teacher Impact

Based on their study, Nye et al. (2004) indicated that "there are substantial differences among teachers in the ability to produce achievement gains in their students" (p. 253). The study was not successful, however, in identifying what specific teacher characteristics were responsible for the increased achievement. It was also suggested that teacher quality and the amount of time dedicated to direct instruction outweighed the students' SES as the primary predictor of academic growth. Sanders and Rivers (1996) had argued in a previous study that with appropriate measurements of teacher effectiveness, administrators have leverage.

> . . . undeniable opportunities to minimize the near permanent retardation of academic achievement of many students resulting from experiencing the most hurtful teacher sequences. If the magnitude of the cumulative effects is not diminished, then students are de facto being placed involuntarily in a lottery where the "luck-of-the-draw" of the teacher sequence may play a most important role in their life's opportunities. (Sanders & Rivers, 1996, p. 6)

If the assertion is accepted that teachers do indeed wield great power over student learning, then should they not be held accountable for that learning? Legislatures, governors, and even the president of the United States believe that they should.

Metrics of Accountability

The goal is to find a way to measure and quantify what matters most in the relationship between human beings as they interact with content, knowledge, and skills. A common statement in business cultures is “What gets measured gets done.” With the trend toward employing business practices to increase accountability for student learning, it is imperative that the metrics that are identified as indicators of achievement, academic growth, and teacher quality are valid and measure what they are intended to measure (Hess & Fullerton, 2010). It is also a necessity that the metrics are reliable and consistent in that measurement. Statisticians along with policy makers continually search to find metrics that can be applied in a practical, efficient, and cost effective manner (Hess & Fullerton, 2010).

In Ravitch’s (2010) commentary on the effects of the Bush and Obama administrations’ policies affecting schools, she emphasized the public’s frustration with education and the unrelenting move to do something dramatic to reverse the trend. Billions of dollars through the Race to the Top program are riding on the dramatic and innovative models that have been identified by the Washington bureaucrats and wealthy foundations: Gates, Walton, Dell, Broad, Fischer, Kellogg, and Lilly (Hess & Fullerton, 2010). The narrative regarding the innovative models that tie test scores to teacher effectiveness is forceful and confident. States are lining up to create teacher evaluation instruments with little or no data that supports the relationship between teacher quality and standardized test scores (Ravitch, 2010). Public dissatisfaction and a sense of urgency create the conditions for a “one size fits all” solution. Hess and Fullerton (2010) argued, “A focus on collecting student achievement data in the past 10 years has increased the amount of information school districts have, but these data are insufficient” (p. 1). Hess and Fullerton suggested that integrating metrics identified by successful business models could provide a menu of measurements that would be an effective tool for improving education. Typically effective business models attribute their success to a system of rewarding positive contributions and weeding out substandard performance. This organizational philosophy is now firmly entrenched in the present national educational reform movement, although a significant number of reformers warn that using monetary rewards at best is ineffective and at worst does harm in terms of increasing positive performance levels (Pink, 2009). Ingvarson and Rowe (2007) reiterated Hess and Fullertons's assertion for the most part. Ingvarson and Rowe (2007) argued,

> There is a crucial need for both a substantive teacher-quality/student-performance/merit-pay research and policy agenda to one that focuses on the need for capacity building in teacher professionalism (and its evaluation) in terms of teaching standards related to *what teachers should know and be able to do.* (p. 1)

Rewarding performance. In an effort to reward effective teachers for their contributions to students' learning and motivate others to become better, a majority of the states are relying on a corporate management model to reform their teacher evaluation practices. Most models include student academic performance data (Mendro, 1998). States like Louisiana and Tennessee have adopted a complex value-added model that includes critical individual student factors such as SES, special education diagnosis, attendance, discipline history, as well as a three-year history of test scores in core subjects. Because of the statistical complexity, the value-added models have a number of significant limitations and most states have chosen a less complex technique to derive student achievement metrics (Kane & Staiger, 2008; Sanders, 1998).

In the case of Indiana, the accountability process is matched more closely to the Colorado, Delaware, and Florida models. The process is intended to generate an estimate of the contributions from all teachers, even those who do not teach subjects or grades that are not annually tested. Achievement data as well as a calculated *growth score* is now included in assigning a grade to each Indiana school (Indiana Department of Education, 2011a). However, the reality of accessing valid data is almost impossible. The tested grades and subjects used by most states are those required by NCLB: math and reading in Grades 3-8. Linking the scores of students to teachers in those grades and subjects is less difficult but still imperfect (McCaffrey, Lockwood, Koretz, & Hamilton, 2003).

During the session of 2011, the Indiana legislature codified the use of standardized test scores as a component of a teacher's evaluation. Senate Enrolled Act I (SEA I; Indiana Department of Education, 2011d) specified that 51% of the teacher's evaluation would be based on student achievement scores, and teachers who were rated ineffective or improvement necessary would not receive any increase in pay. The law also stated that schools should avoid putting a student in a classroom taught by an ineffective teacher for two consecutive years. As a direct result of SEA I, Indiana developed and recommended that school districts adopt the RISE evaluation and development system to ensure teacher quality (Indiana Department of Education, 2011b). RISE is based on the following premise:

> The teacher's Primary Objective accounts for the learning of all students in a class and all content standards in a course. Whether a teacher earns a Highly Effective, Effective, Improvement Necessary, or Ineffective rating depends on the extent to which he or she moves students from their starting point to achieve mastery. (Indiana Department of Education, 2011b, p. 7)

There are two significant components of the RISE evaluation system: professional practice and student achievement. The professional practice combines evidence of instructional knowledge and skills, planning, and leadership with student academic progress measured in three ways: individual growth model data, a school-wide learning measure, and student learning objectives. Teachers can be rated as *highly effective, effective,* or *improvement necessary*. The summative calculation to determine a teacher's rating is based on four principles:

1. Teachers should be treated as similaly as possible.
2. Classes that aren't covered by growth-model data should not be excluded or drastically underrepresented in the final weighting.
3. A teacher's mix of growth model and non-growth model classes should be reflected in the calculation.
4. Data in which we have most confidence is given the most weight. (Indiana Department of Education, 2011a, slide 9)

It is clear that parents, teachers, and the community in Indiana and other states are keenly aware of the importance of an educated populace. Even if the motivation for accountability might be in most cases economic, the common good and the welfare of each and every child who attends a public school cannot be underestimated. However, as simple and obvious as that sounds the reality of what is measured and how it is measured depends upon which constituency is utilizing the data (Goe & Holdheide, 2011).

Presently, there are very different camps, contradictory courses of action, and conflicting expectations of what constitutes an educated student and thus an effective teacher. The competing philosophies tend to align with either the economic business model based on monetary performance incentives that champion "What gets measured gets done" or the broader philosophy that integrates behavioral, social, and neuroscience as significant partners in implementing quality curriculum and instruction that develops *knowledge workers* (Pink, 2009; Wagner, 2008). The business model supports an emphasis on preparing students for standardized tests, setting benchmarks of achievement, and holding kids and teachers accountable. Additionally, many states have adopted the practice that requires that Grade 3 students pass specific reading and math portions of the state standardized test or they are retained in Grade 3. Kansas, Ohio, Florida, and Indiana are only a few of the states that have mandated retention based specifically on third-grade reading scores (Penfield, 2010).

Complexity and ideology. Whether the focus is economic growth for the community or an educated work force that can compete globally, the ultimate goal is the same. The debate about how to achieve that goal is ideologically driven and at times educators and policy advocates lose sight of where the real learning and intellectual engagement occurs. It happens between and among students and teachers in America's classrooms every day. The interaction is complex; therefore, school accountability and quality indicators of effective teachers are issues that those who advocate policy must consider carefully in terms of whether they support or impede effective teaching and learning (Garii, 2006).

Influential models. The significant studies that have been done to determine teacher impact have relied on standardized test scores obtained from the districts' or states' high-stakes accountability tests. In a design to determine the individual teacher influence on student achievement, Sanders and Rivers (1996), associated with the Tennessee Value-Added System, based their model on Grades 2 through 8. Approximately 3 million student scores in mathematics, reading, language arts, and social studies were available from the Tennessee Comprehensive Assessment Program administrations. A cohort of Grade 2 students was followed for four subsequent years. Teacher effects were estimated from a longitudinal analysis by using a statistical mixed model process. Researchers found that there were dramatic differences in achievement gains of students who were taught by a series of highly effective teachers compared to students who were taught by a series of ineffective teachers. With second grade scores equalized for an even start, the teacher sequences resulted in a mean range difference of 52 points in Grade 5 (Sanders & Rivers, 1996).

In Texas, studies have determined similar outcomes. Jordan and Mendro (1997) found in working with the Dallas Independent School System that Grade 4 students who were assigned highly effective teachers in sequence increased from the 59th percentile to the 76th percentile. The groups of students who were assigned a sequence of three ineffective teachers decreased from the 60th percentile to the 42nd percentile (Jordan & Mendro, 1997).

To create a clear understanding of the pressures of accountability and the implications in terms of teacher quality, it is important to examine the influence of the teachers in the primary grades. Standardized testing typically begins at the Grade 3 level. In Indiana, the ISTEP+ test is administered in late April of the student's third-grade year (Indiana Department of Education, 2011c). For all intents and purposes, the Grade 3 teacher assumes the responsibility for the scores on that test. It is evident that the Grade 3 teacher does not have the sole impact on a student's achievement; therefore, it would be useful to determine the educational contribution of the student's previous teachers. The issues of teacher impact on student learning outcomes have focused primarily on Grades 3 through 8; however, less focus has been directed toward the probable cumulative impact of teachers over time within a school (Bressoux & Bianco, 2004; Kane & Staiger, 2008; McCaffrey et al., 2003). In fact, there is scant research available that provides valid measures to determine and, thus, quantify the impact of teachers who are not specifically linked to students in the grades assessed by the standardized high-stakes test.

Unmeasured Contributions to Learning

How is student academic growth measured in the grades prior to the official standardized achievement test? Teachers bring with them to the classroom prerequisites such as content knowledge, experience, coursework, and verbal ability. Each of these elements impacts student achievement (Marzano, 2010). Children bring a variety of skills, experiences, knowledge, and abilities to their educational experience as well. Bressoux and Bianco (2004) noted that there was a gap in the literature regarding the possible cumulative effect. They attributed this gap to the inability to make reliable student–teacher links over time. Presently only about 30% of the states have the capacity to make the links that would be valid in holding teachers accountable for student achievement (Berry, 2007). In any one school there could be varied combinations of teachers, co-teaching configurations, and issues associated with climate and culture that could significantly impact individual student learning and academic success (Aaronson, Barrow, & Sanders, 2007; Marzano, 2003, 2010).

Research suggests that dropouts arise from an accumulation of various risk factors throughout children's schooling that peak in high school (American Psychological Association, 2011). It is increasingly evident that school dropout prevention must begin as early as possible. Some researchers have identified early predictors of dropout in children before they are enrolled in kindergarten. Vocabulary development appears to be a correlating factor in school failure. Hart and Risley (1995) found that children living in professional homes heard 382 words an hour whereas children raised in welfare homes heard an average of 167 words an hour. High performing Grade 3 students had vocabularies about equal to the lowest-performing 12th graders, and high school seniors near the top of their class knew about four times as many words as their lower-performing classmates (Barr, Kamil, & Mosenthal, 1991).

An optimal time in brain development is during the early childhood years between birth and four years old. Unfortunately this is the period when most children are vulnerable to environmental risks associated with poverty. It has been found, however, that quality early childhood education and literacy development are vital and can mitigate the factors that tend to interfere with cognitive and language skills development (Perrone, 1991). There is little argument among scholars as to the value of quality preschool experiences. Studies have indicated that teachers who understand early learning and apply that knowledge to their interactions with children have a positive impact on early childhood education (Gordon, 2006; Woodhead, 2006).

Acknowledging, investigating, and subsequently measuring the contribution and effectiveness of teachers who comprise the non-tested subjects is probably the most challenging aspects of designing a comprehensive and valid teacher evaluation system (Goe & Holdheide, 2011). Presently districts have no standardized tool to measure the impact of ancillary contributions, and there is little literature that provides insight into the individual and cumulative influence that primary teachers have on student learning. Formative assessments that track very specific elements of skills development are widely used throughout the United States. Unfortunately, in the search to identify effective teachers, formative assessments in the early grades often are used to make determinations *of* instruction, rather than *for* instruction. These are determinations that the assessments were not designed to make.

Purpose of Assessment Tools

In Indiana, districts have been provided access to the Dynamic Indicators of Basic Early Literacy Skills (DIBELS, Center on Teaching and Learning, n.d.). The DIBELS measures were specifically designed to assess three of the five Big Ideas of early literacy: phonological awareness, alphabetic principle, and fluency with connected text. DIBELS is a common formative assessment approach that teachers use in determining a child's progress in the curriculum, often called mastery measurement. It is considered a general outcome measure (GOM) whereby a teacher can determine if the student is learning and making progress toward the long-term goal. GOMs are designed to assess the acquisition of a sequence of skills. As students move along the continuum, teachers use the GOM formative assessment to inform their instruction and provide interventions to support struggling students. Formative assessment is characterized as assessment *for* learning and is contrasted with summative assessment, which is assessment or evaluation *of* learning. The ISTEP+ is a summative assessment and Indiana plans to use student scores as a significant measure of teacher effectiveness (Indiana Department of Education, 2011c).

Relying on ISTEP+ as the measure for teacher effectiveness is complex, and conventional wisdom continues to identify a myriad of implications and unintended consequences associated with the use of any one measure that represents teacher quality or student achievement. With that said, at least in Grades 3-8, states have access to summative assessments. The necessity to investigate the individual or cumulative teacher contributions to student learning is increasingly more important. The academic foundation derived from the primary grades must be considered. A typical design utilized in testing students in Grades 3-8 was not designed for and is incapable of providing useful data regarding early learning. Issues relate to the tests themselves give reason for caution and include the probable negative impact on student curiosity and experimentation; the distortion of curriculum, teaching, and learning; and the lowering of expectations.

Tests that are appropriate for Grades 1 and 2 are different from those used in Grades 3-6. Tests for younger children depend on pictures and vocabulary and administrator interpretations, but later tests place greater stress on content. Consequently, high scores in early testing may not carry over to later testing (Perrone, 1991). Assessments designed as formative measures cannot be and must not be substituted to measure the foundational learning. On the other hand, it is imperative that the contributions that primary teachers bring to student achievement be explored and quantified (Meisels, 2003). The limitations identified in assessing young children complicate the ability to capture the desired teacher effectiveness in the primary grades.

Hanushek (2010) supported the notion that teachers add the most value to improve student achievement. In his research analysis of compensation models based on the value added by individual teachers, he remarked that

> cataloging the potential imperfections of value-added measures is simple, but so is cataloguing the imperfections of the current system. . . . Even flawed value added measures could advance the current system of personnel decisions that relies on limited information about teacher effectiveness and often provides weak performance incentives to teachers and administrators. (Hanushek, 2010, p. 9)

In essence, Hanushek (2010) posited that it would be prudent to implement new systems if our intent is to use student data to reward effective teachers and motivate others to get better. What we typically use as evaluation of effectiveness will not work because performance metrics are not a part of the process.

The failure of teacher evaluation systems to provide accurate and reliable information about effective instructional practice continues to confound school districts across the nation. Most school districts operate under the assumption that teachers are interchangeable parts of the education mechanism. Very often the teacher transfer policy allows for or even requires a teacher licensed in a particular discipline or for a range of grade levels to move from school-to-school or grade-to-grade without taking into consideration how one teacher differs from another in terms of strengths and weaknesses. *The Widget Effect* (Weisberg et al., 2009) described this pervasive systemic model that fails to recognize and respond to variations of teacher effectiveness with kindergarteners as opposed to fifth graders, students from impoverished backgrounds from those of affluence, or children in a rural setting from those in urban areas. Teacher instructional practice and social interaction within the classroom are reduced to a set of behaviors that can be identified with a checklist (Weisberg et al., 2009). Ingrained behaviors and district policies exacerbate the problem. As a rule teacher evaluation systems do not differentiate among the most effective teachers to the least effective teachers.

Summary

There is much to be fixed in public education. Teacher quality and classroom practice are now under the microscope. As research interfaces with pedagogy, educators must be cognizant of the limitations of the commingling as well as the obvious benefits. Research can inform our classroom practice, but we should not expect research to provide the silver bullets. A collaboration or integration between research and pedagogy can ensure that decisions are informed by the facts and insights that science can provide. Because teaching and learning are not isolated and insulated interactions among human beings, education leaders should not expect research to ultimately resolve disparate policy disputes. Research and data can, however, create discontent with present policy and practice. It can create a climate of continual questioning and analysis that can be based on more than intuition and theoretical suppositions. The thoughtful assimilation of data and research into educational policy issues provides powerful levers for improving schools and student learning (Hess, 2009).

On any given day, the state of education is either at the forefront of the public debate as with the resignation of Michele Rhee (Moroney & Young, 2010) and the 2010 premier of filmmaker Davis Guggenheim's *Waiting for Superman* (Corliss, 2010) or a continual simmering undercurrent. Public scrutiny and accountability for student achievement are not going to diminish; therefore, deliberate and thoughtful consideration of the commonly used indicators of teacher quality must closely relate to achievement gain. Presently, observed teacher characteristics do not necessarily represent teacher quality. It is also unlikely that a single standardized test score will give a much better measure of effective teaching (Hanushek, 2010). The review of the literature suggests that in trying to improve education in America, the general public, policymakers, educators, philanthropic enties, and social scientists continue to uncover and subsequently try to explain and measure. Teaching and learning continues to be a complex set of variables that are associated with a phenomenon that is uniquely personal and therefore elusive. The present reality in public school is that the chance of a student succeeding academically might be determined on the luck of the draw.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

The review of the literature underscored the significant interest in and resources devoted to quantifying the impact specific teachers have on student achievement in school as measured by standardized achievement test scores. In the age of accountability the mantra increasingly has become what gets measured gets done. The distinct reliance on the business model and econometric research have exacerbated the notion that numerical values can be assigned to just about any endeavor if the desired result is identified in a clear and concise manner. For every X there can be calculated a corresponding Y. If the predicted Y is not achieved, then the X is faulty. In terms of learning that would translate to for every day of classroom instruction by teacher X, there will be a quantifiable amount of learning by student Y. Value-added accountability models attempt to control for effects such as SES, gender, teacher experience, and many other factors that appear to impact student achievement.

This equation is not new to the educational community. Issues around student SES, ethnicity, and gender had previously focused the debate in terms attributed to student deficiencies. If insufficient learning occurred in Y, then it was identified as a student deficit (Coleman et al., 1966). Although educational policy had already begun to demand more accountability from districts, administrators, and especially teachers, a more aggressive shift toward the significance of quality instruction occurred after the signing of the bipartisan No

Child Left Behind federal law in 2002. All students, no matter what their circumstances, were expected to achieve AYP as measured by the state's standardized yearly test.

Linking the findings from the social sciences and educational research, leading contributors and education experts began to identify effective instructional practices and indicators of teacher quality (Marzano et al., 2001; Sousa, 2010). In an effort to account for and mitigate the impact on standardized scores attributed to SES and equity of opportunity, economists, researchers, and statisticians moved to create innovative value-added statistical models (Gordon, 2006; Hanushek, 2010; Sass, 2008; Wong, 2004). Ingvarson and Rowe (2007) noted,

> Methodological limitations endemic to econometric research focusing on the link between teacher quality and student academic performance is well established. An extensive body of work indicates that the typical single-level econometic models fail to conceptualize, measure and evaluate teacher quality in terms of what teacher know and do. Additionally models rarely account for the mesurement, distributional and structural properties of the data that frequently yield misleading interpretations of findings. Findings from multilevel models to relevant data identify variation in students' learning and achievement progress are at the class/teacher level, and have the potential to assist in specifying and evaluating teacher quality in terms of what quality teachers know and are able to do. (p. 2)

Since there is a growing body of work supporting the conclusion that the teacher is the most powerful influence on student achievement, identifying the characteristics of effective teachers and quantifying the attributes associated with teacher quality are critical (Anderson, 1979; Darling-Hammond, 1997; Hattie, 2008; Haycock, 1998; Marzano et al., 2001). There are substantial implications in terms of education policy, political power, and resource distribution. The review of the literature established the widely used link between standardized test scores to school attainment, earnings, and aggregate economic outcomes (Angus & Mirel, 1999; Hanushek et al., 2008). As stated previously, the purpose of the study was to explore the impact that first-, second-, and third-grade teachers had on Grade 3 ISTEP+ scores. In doing so, the study would add further insight into the national accountability movement and increase the body of relevant research regarding the teacher's influence on student learning. The objective of the study was exploratory with no preconceptions regarding best practice instruction. There was no intent to link quality or quanitity of student learning to any particular teacher practice. Quantitative analysis of existing Evansville Vanderburgh School Corporation ISTEP+ data were used to explore teacher impact on student performance.

Problem and Purpose Overview

During the spring, the Indiana State Test for Educational Progress (ISTEP+) is administered to students in Grades 3 through 8 throughout the state. This one assessment measure has profound impact on school accountability and accreditation, resource allocation, economic investment, and, since the passage of SEA 1, teacher evaluation and compensation (Indiana Department of Education, 2011c). The present system is designed to attribute and assign third-grade achievement scores to third-grade teachers, fourth-grade scores to fourth-grade teachers, and so on. The baseline for the student growth score is set at Grade 3; therefore, it is implicit that the prior three teachers have had little impact on student academic foundations and likelihood for future achievement. The state model does not account for students' previous academic experiences in lower-elementary grades, which cannot be reliably tested with paper-and-pencil or computerized tests (i.e., Grades K-2). Likewise non-tested courses such as art, music, drama, or industrial technology that clearly support increased academic achievement cannot be adequately or completely measured (Perrone, 1991). The purpose of this study was to explore and quantify the contributions to student achievement from previous teachers on Grade 3 ISTEP+ scores.

The Research Questions

The research questions in this study employed both descriptive and inferential statistical analysis.

1. Are there recognizable patterns of success in first-, second-, and third-grade teacher combinations in third-grade ISTEP+ scores? ISTEP+ scale scores were organized and summarized by using a descriptive model which included the following: the mean, standard deviation, and standard error for combined math and ELA of the cohort; math and ELA separately for the cohort as a whole; the mean, standard deviation, and standard error for each school, grade level, and classroom separately.
2. Are there significant differences in classroom mean ISTEP+ test scores among teachers at an individual grade level? This question was addressed with an ANOVA that used the *F*-ratio to test for classroom differences for each teacher at each school at each grade level. For example, would the mean difference in the third-grade ISTEP+ score be significant if a student had been in teacher A's first-grade class as opposed to teacher B's first-grade class?
3. Are there patterns of success in first-, second-, and third-grade teacher combinations reflected in third-grade ISTEP+ scores? Do first-grade, second-grade, and third-grade teacher combinations predict a significant proportion of the variance in third-grade ISTEP+ scores? The question was addressed by using a simultaneous multiple

regression model. By utilizing a dichotomous coding system, the exposures to the unique teacher combinations were analyzed. For example, is there a significant difference in scores for students who have the teacher combination X as opposed to students who have the combination Y?

4. Are third-grade ISTEP+ scores predictable based on first-, second-, or third-grade teacher? This question was addressed using a simultaneous multiple regression model. Dichotomous-coded variables Y or N identified a student's exposure to a particular teacher. An equation that provided the line that best described the relationship for the set of X and Y data were generated. Using the defined equation, a relationship between teacher exposure and student ISTEP+ score was determined then analyzed.

Null Hypotheses

$H_0$1. There are no recognizable patterns of success on grade ISTEP+ scores that can be attributed to any single or combination of students' first-, second-, and third-grade teachers.

$H_0$2. There are no significant differences in third-grade classroom mean ISTEP+ scale scores among students assigned to specific teachers at any previous individual grade level.

$H_0$3. There are no patterns of success in first-, second-, and third-grade teacher combination reflected in third-grade ISTEP+ scores. A significant proportion of the variance in third-grade ISTEP+ scores is not predictable using first-grade, second-grade, and third-grade teacher combinations.

$H_0$4. Third grade ISTEP+ scores cannot be predicted based on first-, second-, or third-grade teacher.

Population

I am a district level director of schools including K-8, K-5, K-6, Grades 6-8, and Grades 7-8 configurations in EVSC. Duties associated with teacher staffing, curriculum, and assessment fell within the scope of my job description. Access and analysis of student achievement data records were a significant expectation included in my purview.

The study participants included 339 children from seven elementary schools in a suburban/urban district (EVSC) located in southern Indiana. The population make-up of EVSC included a relatively stable cohort of students in each school to study. The students' demographics reflected the neighborhoods in which they live, ranging from affluence to poverty; stable family units to chaotic living arrangements. The elementary schools were selected based on the stability of grades taught by teaching staff in Grades 1 through Grades 3. The teaching assignment for the identified teachers remained constant for the schools years 2008-2009, 2009-2010, and 2010-2011: the primary grade teachers had taught the same grade level during the three years designated for the study. It was determined that three grade levels would provide a sufficient sample size for the analysis. Kindergarten was eliminated as part of the study. It was concluded that four grade levels would limit the potential for a sample size that would produce usable data. Twenty schools were potential participants. Schools with at least two classrooms per grade level were selected. Thirteen schools were excluded due to the instability of teacher grade-level assignments over the three-year time frame being studied. The Grade 1 students in 2008-2009 were designated as the cohort to be studied in each of the elementary schools. The population consisted of third-grade students in 2010-2011 who had attended the same school from first grade through third grade and who were administered the ISTEP+ test at the end of

Grade 3. Students who changed classrooms or moved schools were excluded even if they returned.

The study was structured using students who, as of October 1, established a valid link with a teacher for each school year; 2008-2009, 2009-2010, and 2010-2011. Student–teacher matches were considered valid if the following criteria were met: The teacher taught a *self-contained* (i.e., all day, all subjects) class for the relevant grade and year; the class was not designated as special education or high ability class.

Each school studied was described as either predominantly suburban or urban. Schools tended to reflect the demographics of the neighborhood, and because students are not randomly assigned to schools, data from each school were described and analyzed separately.
During the scope of the study, however, some students were enrolled in schools that were out of their neighborhoods. NCLB provides parents the opportunity to choose other public schools or take advantage of free tutoring if their child attends a school that needs improvement (U.S. Department of Education, 2012a). Students are randomly assigned to classrooms.

Type of Study and Instrument

The model of inquiry utilized existing data from the 2010-2011 school year. By using information through mathematical procedures (statistics), the intention was to inform decisions regarding curriculum and instruction, teacher quality, and the allocation of resources. The achievement scores for ELA and math ISTEP+ tests were the basis for the quantitative analysis. For the purposes of educational accountability, the Indiana Department of Education uses several measures provided by the ISTEP+ test to report school and district success in terms of student achievement: student achievement criteria referenced scale cut scores based on the Indiana academic standards for each grade and subject matter and student growth percentiles based on an individual's test performance relative to his peers in math and ELA from year to year. The baseline or growth cohort is established using the student's ISTEP+ scores in Grade 3. Within the two larger measures, the state examines the performance of subgroups in academic achievement and academic growth. The state accountability model uses an A through F grading system: A = 4, B = 3, C = 2, D = 1, F = 0. The school grade is based on the achievement measure and scale scores. The growth measure is used to determine whether the school receives additional points or whether points are subtracted in accordance with the performance of the subgroups in both academic achievement and academic growth.

This study utilized ISTEP+ data that measured student achievement levels in English/language arts, mathematics, science, and social studies. The data contained test scores for all Indiana students attending public and non-public schools in Grades 3 to Grade 8. End-of-course assessment (ECA) scores were collected for Indiana students in Grades 9 and 10 as well. The data also include information on the schools' location and demographics: ethnicity, gender special education classification, and eligibility for free or reduced lunches (SES). ISTEP+ reports student achievement levels according to the *Indiana Academic Standards* that were adopted in November 2000 by the Indiana State Board of Education. The instrument itself was developed to align with achievement level descriptors (ALDs) that correspond to the knowledge, skills, and abilities found in the Indiana Standards. ALDs are commonly used in the testing community to describe test scales and communicate performance across achievement levels on the test scale. As a mandated component of NCLB, states must provide peer-reviewed descriptions of the competencies associated with an achievement level. A cut score that represents minimal competency is set for each ALD. As an additional requirement of NCLB, ISTEP+ also provides three cut points that indicates a score that falls below proficient, proficient, and above proficient. For the purposes of this study, the third-grade scale score that represent minimal competency was utilized in the data analysis.

The Variables

The independent variables in this study were the teacher links. The student's ISTEP+ score was the dependent variable. The score was linked to the first-grade teacher, the second-grade teacher, and third-grade teacher. The score was also linked to combinations of grade-level teachers (teacher chains). For example, Student A had Teacher 1a for first grade, Teacher 2d for second grade, and Teacher 3g for third grade. Student B had Teacher 1a for first grade, Teacher 2e for second grade, and Teacher 3h for third grade. Student A's individual teacher links (1a, 2d, 3g) were correlated to her third-grade ISTEP+ score. Student B's score were also correlated to each of the multiple combinations of links (1a, 2e; 1a, 3h; 2e, 3h; 1a, 2e, 3h) and analyzed. The multivariate data were linked to the dependent variable, Grade 3 ISTEP+ scale scores in ELA and math.

Data Analysis

ISTEP+ math and ELA scale scores were examined from a sample ($N = 339$) derived from a pool of 768 students enrolled in an EVSC first-grade classroom in October 2008 in one of five elementary schools. The students who were excluded from the study did not meet at least one of the expectations criteria—for example, if the student moved from the school or changed classroom teacher at some point during the length of the study.

Participants

Students in each of the seven elementary schools were placed in first-, second-, and third-grade classrooms based the process defined in Chapter 1 as random distribution. The population parameters were third-grade scores, first graders in 2008-2009, and students who attended one of the five identified elementary schools and remained in that school throughout the three-year time frame studied. To provide more information about the study participants, descriptive data were generated to answer Question 1. Descriptors including the mean achievement score and standard deviation of the study population's third-grade ISTEP+ scores in math and ELA were generated in total, by school, and by grade level and individual teacher. To address Question 2 regarding the impact of teachers at an individual grade level, data provided by an ANOVA tested for any classroom differences for each teacher at each school at each grade level.

The impact on the variance of student scores was determined statistically by explaining the percent of variance explained. A basic assumption underlying the use of this index was that the percentage of variance explained by a predictor (independent) variable (e.g., teacher) relative to a predicted (dependent) variable (e.g., student achievement score) represented the strength of relation between the two. Most commonly a set of predictor variables was used.

For example, this study sought to predict student achievement using (a) first-grade teacher, (b) second-grade teacher, and (c) third-grade teacher. These predictor variables accounted for a proportion of total variance in the third-grade ISTEP+ score.

The process of determining the relationship between a predicted (dependent) variable and a predictor (independent) variable is commonly referred to as regression analysis. A regression analysis was used to provide data to address Question 3. Individual teachers and combinations of teachers were coded as dichotomous variables. Simultaneous multiple regressions generated data to analyze the possible impact of the exposure to a unique teacher or a combination of teachers (teacher chain) on student achievement. The total impact on student achievement was investigated by using regression analysis. The analysis was designed to explore the proportion of impact that might be attributed to a single teacher.

A second regression analysis was run using the combination of a set of teachers as predictors of student achievement. Multiple predictors (teacher combinations or teacher chains) were used with a predicted variable; the relationship between the predictor variables considered as a set, and the predicted variable was expressed as *R*, the multiple correlation coefficients.

Finally, Question 4 was addressed using a regression analysis to determine if ISTEP+ could be predicted from any one individual grade level teacher. A single predictor or independent variable (teacher) was used with a predicted or dependent variable (student achievement), the relationship between the two was expressed as *r*, the Pearson product moment correlation.

Summary

In this chapter, the hypotheses, data source including population, and the analyses used were presented and described. The main purpose of the study was to explore and quantify the contributions of previous teachers to student achievement that may be detected utilizing Grade 3 ISTEP+ scores. In Chapter 4, a broad analysis of existing ISTEP+ data shows if any effect on student achievement was derived from students' exposures to individual teachers or combinations of teachers in grades prior to the administration of ISTEP+ in the third grade. The investigation included descriptive analysis, ANOVA, and regression analysis. In Chapter 5, the results and implications of the study are reported and suggestions for further study are discussed.

CHAPTER 4

RESULTS

The data analysis and results obtained from the statistical models were designed to explore and quantify the contributions of previous teachers to student achievement among the study sample. Chapter 4 provides the analysis to determine the impact of the exposure to any one teacher or combination of teachers to student achievement as measured by third-grade ISTEP+ scale scores in both math and ELA. The statistical results were used as a non-experimental method for estimating teacher effect on student test scores.

Chapter 4 begins with a description of the data collected and the criteria used to select the schools and student samples for this study. Included in the description is an overview of each school's demographic characteristics separately and in relationship to the larger district's demographics. The overview also included an explanation of the unique identifiers used to link school, teacher, and student data to prepare them for the statistical analysis.

Chapter 4 explains the statistical models used in the study beginning with the method employed to create the data set. The EVSC schools that were selected met the primary condition that the first-, second-, and third-grade teachers in the school taught the same grade level all three years of the study. Within the seven schools, a pool of 768 third-grade students who were administered the spring assessment in 2010-2011 and received both ELA and math ISTEP+ scores served as the basis for the analysis set logic. The pool was narrowed to a sample cohort of

339 students by eliminating students who were not in the same school for three years; students who did not have both ELA and math ISTEP+ scores; students who were not first graders in 2008-2009, second graders in 2009-2010, and third graders in 2010-2011; and students who were linked to teachers who had changed grade levels or transferred from the school during the time of the study.

Each of three subsequent statistical models provided data for analysis with the purpose of addressing the three research questions. The first model provided descriptive statistics: number of participants and measures of central tendency based on ELA and math ISTEP+ scale scores. The second model included an ANOVA based on the variance of the ISTEP+ scale scores within each school and grade-level classroom for both ELA and math. The third model featured a multiple regression that estimated the proportion of variance attributed to unique teacher combinations in student ISTEP+ math scale scores and ISTEP+ ELA scale scores. A series of independent variables derived from unique teacher combinations within each school was utilized. The fourth model utilized a multiple regression model as well. Using the defined equation, the model estimated the relationship between dependent variables, ISTEP+ math and ELA scale scores, and the exposure to the independent variables of a specific teacher link.

The targeted population for this study was derived from EVSC district data that included students during the study period 2008-2009 through 2010-2011. A total of seven schools were selected from a pool of 20 elementary schools. The criteria used to narrow the data set were introduced and filters were set to include only third-grade ISTEP+ math and ELA scores for a cohort of students that met the following conditions:

1. Students were first-grade students in the EVSC and included in official enrollment October 1, 2008.

2. Students remained in that particular school throughout the following three-year period and did not withdraw or reenter at any time.
3. October 1 of each year, the student was linked to a teacher and remained in that teacher's classroom throughout the entire school year.

The narrowed sample was audited and examined under the assumptions that students had been randomly assigned to teachers each of the three years of the study and classroom configurations were proportionate with the general population percentage in terms of special needs or high-ability students.

For the purposes of this study, the schools were identified as school A, B, C, D, E, F, and G. Teachers were identified in the following manner: School A, teacher 3n was a specific third-grade teacher in school A. The other teachers in that grade were identified as 3o, 3p, and 3q. First-grade teachers in school A were identified as 1n, 1o, 1p, and 1q. Each school and teacher identifiers followed the same pattern throughout Chapters 4 and 5.

1. *Are there recognizable patterns of success in first-, second-, and third-grade teacher combinations in third-grade ISTEP+ scores?* This section discusses the results of Question 1 in descriptive statistics including the mean, standard deviation, range, skew, and kurtosis. District and school demographics were also included to provide context.

The EVSC is the third largest school district in Indiana with an enrollment of about 22,250 students. The district is predominantly urban, but a large suburban region is situated on the northern and eastern sides of the county. A small rural area is located on the county's northwest side. The student population is 28% minority and 57% of the students qualify for free/reduced lunch. Indiana's demographics are similar with a 28% minority population and 48% of the students living below the poverty line.

In ELA and math, the mean ISTEP+ scale score was below that of the state. The mean ELA 2010-2011 ISTEP+ third-grade scale score for the state (N = 73,586) was 463.33 compared to the EVSC's, which was slightly below at 454.83 (N = 1,664). The math 2010-2011 ISTEP+ third-grade mean scale score for Indiana (N = 73,972) was 470.36 compared to EVSC's mean of 457.81 (N = 1,627). The study sample cohort appeared to be representative of the EVSC population taking into account sampling error.

School A had an enrollment of 476 and was one of nine EVSC schools that were designated as Title I. School A was 60% minority and 88% of the students fell below the poverty line. The school is situated on the southeast side of Evansville and the demographics of the school reflected those of the neighborhood. The school included a pre-kindergarten center that was staffed and directed by the Center for School, Community, and Family Partnerships, a department within the EVSC. The center oversaw several pre-K programs throughout the district and operates collaboratively with the school. School A was staffed with approximately four sections of each grade level, K-5.

The third-grade ISTEP+ ELA scores for School A (N = 38) had a range of 345 points with a mean of 421.61 (SD = 84.81). The distribution differed slightly from normal (*skew* = -1.24; *kurtosis* = 0.85). In math, School A had a range of 440 points with mean of 416.49 (SD = 101.27). The distribution did not differ significantly from normal (*skew* = -.80; kurtosis = 0.21). School A's mean scale scores in both ELA and math fell below the EVSC and Indiana scale scores in 2010-2011. Table 1 contains the information.

Table 1

School A Mean ISTEP+ ELA and Math Scores by Grade Level

School A	Grade 1	Grade 2	Grade 3
Mean ELA (421.61)			
n	426.07	385.64*	427.36
o	424.00	455.77	418.67
p	395.43	411.14	430.10
q	431.09	435.14	409.73
Mean Math (416.79)			
n	419.80	406.09	425.64
o	432.40	459.54*	391.17
p	402.29	378.14	426.60
q	414.82	392.86	413.00

Note. Highlighted scores are > school mean; **SD* ± .40

School B was reconfigured and opened as a K-8 school in 2010-2011. The school had an enrollment of 558 students; 35% were minority and 92% lived in poverty. The school is identified as the heart of a strong community identity and includes a pre-K center. Most students live in close proximity and many typically walk to school.

The third-grade ISTEP+ ELA scores for School B ($N = 26$) had a range of 234 points with a mean of 414.42 ($SD = 57.93$). The scores appeared to be normally distributed (*skew* = -0.58; *kurtosis* = 1.15). In math, School B had a range of 306 points with a mean of 398.46 ($SD = 87.78$). The distribution did not differ significantly from normal (*skew* = -0.50; *kurtosis* =

-0.70). School B's mean scale scores in both ELA and math fell below EVSC by 40.41 points and 59.35 points respectively and below Indiana scale scores in 2010 -2011 by 71.90 points. These figures are available in Table 2.

Table 2

School B Mean ISTEP+ ELA and Math Scores by Grade Level

School B	Grade 1	Grade 2	Grade 3
Mean ELA (414.42)			
n	389.11	406.86	400.64
o	459.29*	427.80	417.33
p	405.80	405.44	435.33
Mean Math (398.46)			
n	372.22	405.71	398.45
o	445.00*	397.70	403.89
p	389.50	393.67	390.33

Note. Highlighted scores are > school mean; **SD* ± .40

School C had an enrollment of 483 students in grades K-5. The school population was 36% minority and 71% of the students were eligible for free/reduced lunches. The school is situated on the near east side of Evansville in an area populated by well-established middle- to upper-income families; however, the school population was predominantly high poverty. Even though School C had a high percentage of students living in poverty, it did not qualify for Title I services because many children with means attended parochial school. School C was staffed with approximately four sections of each grade level, K-5.

The third-grade ISTEP+ ELA scores for School C (*N* = 33) had a range of 256 points with a mean of 463.79 (*SD* = 59.04). The scores did not appear to differ from a normal distribution (*skew* = 0.15; *kurtosis* = -0.12). In math, School C had a range of 363 points with a mean of 492.52 (*SD* = 91.66). The distribution did not differ significantly from normal (*skew* = 0.81; *kurtosis* = 0.39). School C's mean scale score in ELA was above EVSC's and similar to Indiana scale scores. However, in math, the School C mean scale score was above EVSC and the state by 34.71 and 22.16 points respectively. These data are presented in Table 3.

Table 3

School C Mean ISTEP+ ELA and Math Scores by Grade Level

School C	Grade 1	Grade 2	Grade 3
Mean ELA (463.79)			
n	470.33	444.50	469.33
o	455.20	478.44	480.29
p	462.25	453.20	421.29
q		468.90	
Mean Math (463.79)			
n	516.60*	480.75	489.33
o	462.00	511.89*	528.00*
p	485.50	468.70	427.00
q		503.60*	

Note. Highlighted scores are > school mean; **SD* ± .40

School D's enrollment of 848 was 26% minority, and 47% of the students were eligible for free/reduced lunches. The school was situated in a community that is populated predominantly by middle- to upper-income families but also included several low-income areas of housing on the city's east side. School D was staffed with approximately six sections of each grade level K-5 and at one time was the city's largest elementary school.

The third grade ISTEP+ ELA scores for School D (*N* = 63) had a range of 193 points with a mean of 461.95 (*SD* = 42.87). The scores appeared to be a normal distribution (*skew* = -0.34; *kurtosis* = 0.56). In math, School F had a range of 255 points with a mean of 476.0 (*SD* = 56.10). The distribution did not differ significantly from normal (*skew* = -0.62; *kurtosis* = 0.07). School F's mean scale scores in ELA was similar to EVSC's and the state's scale scores but was 20-points higher than EVSC and six points higher than the Indiana scale scores in 2010-2011. Table 4 reflects these findings.

School E had an enrollment of 987 and is 12% minority. The percentage of students eligible for free/reduced lunches was 38%. The school was situated in a predominantly middle-income community on the city's north side. The student population was stable and transience was negligible. School E was staffed with approximately seven sections of each grade level, K-5.

The third-grade ISTEP+ ELA scores for School E (*N* = 101) had a range of 282 points with a mean of 490.27 (*SD* = 49.52). The scores did not appear to differ from the normal distribution (*skew* = -0.21; *kurtosis* = 0.61). In math, School E had a range of 420 points with a mean of 510.05 (*SD* = 82.64). The scores followed normal distribution (*skew* = 0.12; *kurtosis* = 0.68). The mean scale scores of School E in ELA and math exceeded both the EVSC mean and the Indiana mean scale scores. Table 5 presents these findings.

Table 4

School D Mean ISTEP+ ELA and Math Scores by Grade Level

School D	Grade 1	Grade 2	Grade 3
Mean ELA (461.95)			
n	439.00	475.69	420.78
o	447.17	456.67	473.13
p	469.43	462.22	474.06
q	480.22*	455.67	483.38*
r	462.07	459.86	430.78
s	477.00		
Mean Math (476.70)			
n	444.11	484.38	433.00
o	459.83	474.50	489.25
p	448.29	475.22	479.19
q	487.89	483.44	500.08*
r	504.47*	469.00	459.89
s	492.82		

Note. Highlighted scores are > school mean; **SD* ± .40

Table 5

School E Mean ISTEP+ ELA and Math Scores by Grade Level

School E	Grade 1	Grade 2	Grade 3
Mean ELA (494.42)			
n	501.83	494.00	498.92
o	473.15	530.00*	470.57
p	491.75	490.62	489.50
q	487.38	488.69	500.50
r	492.00	480.23	506.88
s	511.57	467.18	496.79
t	508.10	498.83	
Mean Math (517.07)			
n	557.50*	540.58	517.08
o	464.92	535.53	498.14
p	516.44	492.62	525.63
q	494.94	510.00	494.31
r	517.50	489.54	516.65
s	535.21	497.18	543.37
t	546.80	540.83	

Note. Highlighted scores are > school mean; **SD* ± .40

School F had an enrollment of 590 and was 12% minority; 62% of the students were eligible for free/reduced lunches. The school was situated in a community on the city's north side populated by middle-income and working-poor families. School F was staffed with approximately five sections of each grade level, K-5.

The third-grade ISTEP+ ELA scores for School F (*N* = 66) had a range of 269 points with a mean of 461.85 (*SD* = 45.81). The scores differed from the normal distribution (*skew* = -0.62; *kurtosis* = 2.57). In math, School F had a range of 430 points with a mean of 487.14 (*SD* = 73.74). The distribution differed slightly from normal (*skew* = 0.64; *kurtosis* = 1.44). School F's mean scale score in ELA was slightly above EVSC's and within two points of the state scale score. In math School F scores were 29.33 above the EVSC mean and 16.78 above the Indiana mean scale score. These figures are available in Table 6.

School G had an enrollment of 379, was 16% minority, and 74% of the students were eligible for free/reduced lunches. The school was part of an established west-side community populated by middle- to low-income families. The school had very few transient students and was staffed with approximately two sections of each grade level, PK-5.

The third-grade ISTEP+ ELA scores for School G (*N* = 29) had a range of 315 points with a mean of 446.41 (*SD* = 57.47). The scores appeared to differ from the normal distribution gathering near the mean (*skew* = -0.94; *kurtosis* = 4.25). The mean scale score in ELA fell below both EVSC and the state. In math, School G had a range of 261 points with a mean of 438.86 (*SD* = 61.67). The scores followed normal distribution (*skew* = 0.70; *kurtosis* = 0.56). School G scores in math fell below both the EVSC mean and the Indiana mean scale score. These scores are reflected in Table 7.

Table 6

School F Mean ISTEP+ ELA and Math Scores by Grade Level

School F	Grade 1	Grade 2	Grade 3
Mean ELA (461.85)			
n	457.00	461.08	467.53
o	498.00*	451.54	462.36
p	471.89	459.40	437.05
q	457.50	468.13	485.59
r	451.80	469.53	
s	446.46		
Mean Math (487.14)			
n	488.92	475.92	484.07
o	522.78*	473.38	476.43
p	493.89	495.20	460.65
q	479.58	473.50	529.828*
r	483.10	505.53	
s	466.08		

Note. Highlighted scores are > school mean; **SD* ± .40

Table 7

School G Mean ISTEP+ ELA and Math Scores by Grade Level

School G	Grade 1	Grade 2	Grade 3
Mean ELA (446.41)			
n	465.53	440.33	452.00
o	425.93	452.93	442.47
Mean Math (438.86)			
n	456.40	433.60	443.42
o	420.07	444.50	435.65

Note. Highlighted scores are > school mean; **SD* ± .40

Exploring the emergence of pattern in grade-level mean ISTEP+ scale scores by using mean scale scores was problematic. A limitation associated with the model was the possibility that individual students could prove to be overly influential. The review of grade-level scale scores might offer the appearance that the sample mean difference for each individual teacher in each school was more than would be expected by chance. That conclusion necessitated careful consideration. By using the mean measure of central tendency as the metric, individual cases had the ability to significantly affect the mean score for the classroom teacher without providing meaningful data. For example, a student in School A, who obtained a high ISTEP+ scale score in math, might be included in Teacher 2c’s class of very average student scale scores. The mean score would appear to support a significant teacher’s effect on student achievement that was much greater than if the *influential* case had not been included. The decision was made to include cases that were proved to be influential in their impact on the classroom mean score because the assumption was made that each school and each teacher in each grade level had the same chance of having an *influential* score as part of his or her data and, thus, reflected the population parameters. Notable patterns did emerge for further investigation in School C's second-grade math and School D's first-grade ELA. In both schools those particular grade levels suggested that there might have been factors associated with the teacher teams or the grade-level curriculum that would bear further study.

Significance of patterns of success between first- and second-grade and second- and third-grade scores were not considered to be probable because the random assignment of students from first grade to second grade and, subsequently, to third grade negated the opportunity of a measurable correlation. Students had an equal chance of being assigned any teacher in the subsequent grade level.

The analysis used to explore Question 1 suggested little teacher influence regarding first-, second-, and third-grade teachers on student achievement success on ISTEP+. However, it provided a context in which to analyze the findings of the subsequent models.

2. *Are there significant differences in classroom mean ISTEP+ test scores among teachers at an individual grade level?* This question was addressed with an ANOVA that used the *F*-ratio to test for classroom difference for each teacher at each school at each grade level. For example, would the mean difference in the third-grade ISTEP+ score be significant if a student had been in Teacher 1n's first-grade class as opposed to Teacher 1p's first-grade class?

The *F* statistic was calculated in ELA for the cohort (n = 38) in School A. In the three consecutive years of the study, the difference in classroom mean ISTEP+ scores was not significant at the .05 level. The *F* statistic calculated for math for the years of the study indicated no significance at the .05 level, as illustrated in Table 8.

Table 8

Differences Among Teachers Within Each Grade Level – School A

	School Year		
School A	2008-2009 Grade 1	2009-2010 Grade 2	2010-2011 Grade 3
ELA	0.27	1.45	0.12
Math	0.09	1.32	0.18

As provided in the Table 9, the calculated *F* statistic for ELA in School B (n = 26) indicated statistical significance in the first-grade teachers. The ANOVA revealed that at least one teacher in the data set was different than the others. That difference in the mean score was less likely to occur by chance or randomness, $F(23) = 3.74$, $p < .05$.

The *F* statistic identified no significant difference in classroom mean in the study's second and third years at the .05 level. Math mean ISTEP+ scores showed no significant difference at the .05 level.

Table 9

Differences Among Teachers Within Each Grade Level – School B

	School Year		
School B	2008-2009 Grade 1	2009-2010 Grade 2	2010-2011 Grade 3
ELA	3.74*	0.41	0.70
Math	1.50	0.04	0.04

The *F* statistic for School C was calculated in ELA for the cohort ($n = 33$). The data showed no significant difference in classroom mean ISTEP+ scores in ELA at the specified .05 level as illustrated in Table 10.

Table 10

Differences Among Teachers Within Each Grade Level – School C

	School Year		
School C	2008-2009 Grade 1	2009-2010 Grade 2	2010-2011 Grade 3
ELA	0.19	0.44	2.67
Math	1.10	0.41	3.24

The *F* statistic for School D was calculated for the cohort ($n = 63$). The data showed no difference at the specified .05 level among the mean scores linked to first-grade teachers, $F(57) = 1.50$, and second-grade ELA teachers, $F(58) = 0.44$, but did reveal a statistical significance among the mean scores linked to ELA third-grade teachers, $F(58) = 6.25$. In math, the data showed statistical significance at the .05 level in first grade, $F(57) = 2.42$, and in third grade, $F(58) = 2.58$. Data presented in Table 11 showed no difference in mean scores linked to second-grade math teachers, $F(58) = 0.16$.

Table 11

Differences Among Teachers Within Each Grade Level – School D

School D	2008-2009 Grade 1	2009-2010 Grade 2	2010-2011 Grade 3
ELA	1.50	0.44	6.25*
Math	2.40*	0.16	2.58*

Note. * $p < .05$

The *F* statistic for School E was calculated for the cohort ($n = 101$). The data showed a significant difference among the mean scores linked to ELA first-grade teachers at the .05 level, $F(93) = 2.68$, and math first-grade teachers, $F(93) = 3.91$.

The data did not indicate significance at the .05 level for ELA in second grade or third grade. Data presented in Table 12 did not suggest significance for math in second grade or third grade.

Table 12

Differences Among Teachers Within Each Grade Level – School E

School E	2008-2009 Grade 1	2009-2010 Grade 2	2010-2011 Grade 3
ELA	2.68*	2.08	0.73
Math	3.91*	1.14	0.52

The *F* statistic for School F was calculated for a cohort ($n = 66$). The data showed no significant difference at the .05 level among the mean scores in first-grade and second-grade ELA. There was a significant difference in third-grade ELA mean scores, $F(62) = 4.05$, $p < .05$. The data also indicated there was a similar pattern in math.

The data presented in Table 13 indicated no significant difference at the .05 level in first-grade and second-grade math. Statistical significance was revealed in third-grade math, $F(62) = 3.15$, $p < .05$.

Table 13

Differences Among Teachers Within Each Grade Level – School F

	School Year		
School F	2008-2009 Grade 1	2009-2010 Grade 2	2010-2011 Grade 3
ELA	1.74	0.32	4.05*
Math	0.66	0.55	3.15*

The *F* statistic for School G was calculated for the cohort ($n = 29$). The data showed no significant difference in classroom mean ISTEP+ scores in ELA or in math as illustrated in Table 14.

Table 14

Differences Among Teachers Within Each Grade Level – School G

	School Year		
School G	2008-2009 Grade 1	2009-2010 Grade 2	2010-2011 Grade 3
ELA	3.78	0.34	0.19
Math	2.66	0.22	0.11

The ANOVA offered not only data of statistical significance, it also provided an opportunity to add to the discussion regarding tools for school officials and policymakers to include in the pursuit of appropriate measures of teacher performance and its relationship to students' academic achievement. Chapter 5 contains further discussion regarding the practical significance and implications of using the ANOVA model for informing educational decision-making and the placement of teachers in teaching assignments.

3. *Are there patterns of success in first-, second-, and third-grade teacher combinations reflected in third-grade ISTEP+ scores?* The purpose of Question 3 was to look for patterns of academic success by students who have been exposed to specific teacher combinations. A multiple regression analysis was used to provide a statistical description of the relationship between student ELA and math ISTEP+ scores and the cohort's exposure to combinations of teachers in the first-, second-, and third-grades. The intent of the analysis was used to determine whether achievement levels were statistically significant or occurred by chance or randomness. The model offered data that suggested that certain teacher combinations provided evidence of a variance in the test scores worth investigating. The regression computed an F ratio for each

combination of teachers for each school in the study. Because the *F* ratio is a measure of the ratio of systematic variation to unsystematic variation, a large *F* ratio value indicated that the sample mean difference was more than would be expected naturally and would prompt further examination.

The regression design assigned a unique variable to each combination of teachers. For example, in School A, a unique identifier was assigned to students who experienced the teacher combination 1n, 2p, and 3o. Combinations of first and second grade as well as second and third grade were also coded. The coding was binary: either a student had this combination or he or she did not. A school that had four first grades, four second grades, and four third grades had 64 possible teacher combinations. The sample school mean for ISTEP+ math and ELA was used as the constant in the regression equation.

Each combination was loaded by sample school into Minitab, an accepted statistical software program. An output was created that provided an *R*-squared, adjusted *R*-squared, *p*-value, mean and standard error of the mean, *t*-test, and an *F* ratio and within-sample and between-sample degrees of freedom in ELA and math for each sample school. The statistics output offered sufficient data to examine and discuss the hypothesis included in Question 3.

In the process of reviewing the data from each sample school and using the law of large numbers as one factor for analysis, there was reason to suspect that sample size was inadequate to produce meaningful data to examine in first-, second-, and third-grade teacher combination. In many cases the number of students linked to specific teacher combinations included only one student. Because of this limitation, the analysis focused on Grade 2 and Grade 3 teacher combinations for students in each sample school. By examining the R^2 and the adjusted R^2 values for each school, the proportion of variance explained by the model was calculated. R^2 is a standardized statistic that gives a measure of the *goodness of fit* of the estimated regression equation for the sample. The statistic is derived by subtracting the *unexplained* variation in the dependent variable divided by the *total variation* of the dependent variable from one.

$$R^2 = 1 - \frac{\text{SSR}}{\Sigma\,(y_i - \bar{y})^2} = 1 - \frac{\text{SSR}}{\text{SST}}$$

$R^2 = 1$ is a perfect fit and obtained only if the data points lie exactly along a straight line; $R^2 = 0$ indicates that the model is a poor fit and is useless as a predictor. The adjusted R^2 (ΔR^2) is used to measure the equation's *goodness of fit* for the general population. The minitab output provided both statistics for review.

The summary data output for second- and third-grade teacher combinations in School A was not statistically significant in ELA. The regression model predicted that 38% of the variance in student achievement could be explained by the combination of teachers in Grades 2 and 3. When adjusted for sample size and number of predictors, the variance explained dropped to 0%, $R^2 = .38$, $\Delta R^2 = 0$, $F(15, 22) = .90$, $p = .58$. None of the second- and third-grade teacher combinations considered as separate units proved significant at $p < .05$. Upon further examination of the sample, only three of 16 teacher combinations included four or five students at most. The data indicated that 30% of the teacher combinations included one student. In math, 49% of the variance could be explained by the combination; however, when adjusted for sample size, the variance dropped to 14%, $R^2 = .49$, $\Delta R^2 = .14$, $F(15, 22) = 1.43$, $p = .22$. The statistics R^2 and ΔR^2 provided evidence that the sample size might be an influence.

School B summary data also indicated similar findings. The model predicted that only 21% of the variance in ELA student achievement could be explained by the combination of teachers in Grade 2 and Grade 3. When adjusted for sample size and number of predictors, the variance explained dropped to 0%, $R^2 = .21$, $\Delta R^2 = 0$, $F(8, 17) = .56$, $p = .79$. The sample size for each teacher combination was also low; only 22% of the combinations included more than three students. In math, the regression predicted 14% of the variance could be explained by the combination, but when adjusted for sample size the variance explained dropped to 0%, $R^2 = .14$, $\Delta R^2 = 0$, $F(8, 17) = .36$, $p = .93$.

School C summary data showed no significance. The regression model predicted that 27% of the variance in student achievement could be explained by the combination of teachers in Grades 2 and 3. When adjusted for sample size and number of predictors, the variance explained dropped to 0%, $R^2 = .27$, $\Delta R^2 = 0$, $F(9, 23) = .90$, $p = .50$. Fifty percent of the individual teacher combinations included four or five students. Math followed the same pattern with a 37% variance explained by the teacher combinations, but when adjusting for sample size, only 12% could be explained, $R^2 = .37$, $\Delta R^2 = .12$, $F(9, 23) = 1.49$, $p = .21$.

Even though School D had a larger pool of students in the study ($N = 63$), the summary data indicated no significance. The model predicted that 27% of the variance in student achievement could be explained by the combination of teachers in Grade 2 and Grade 3. When adjusted for sample size and number of predictors, the variance explained dropped to 0%, $R^2 = .27$, $\Delta R^2 = 0$, $F(9, 23) = .96$, $p = .50$. Only 29% of the combinations included four or five students. The teacher combinations in math again reflected the same results: 35% of the variance could be explained, however when adjusted for sample size, the variance explained decreased to 0%, $R^2 = .35$, $\Delta R^2 = 0$, $F(23, 39) = .90$, $p = .60$.

School E summary data did not indicate statistical significance for ELA. The regression model predicted a slightly larger variance of 48% in student achievement could be explained by the combination of teachers in Grade 2 and Grade 3. When adjusted for sample size and number of predictors, the variance explained again dropped to 0%, $R^2 = .48$, $\Delta R^2 = .14$, $F(37, 57) = 1.41$, *p* = .12. The data did identify two combinations that suggested further examination might be appropriate. Teacher combination 2o, 3s proved significant, $t(57) = 2.24$, $p = .028$. The number of students linked to the teacher combination, however, was three. The small number suggested that the significance be viewed with caution. The data for teacher combination 2s, 3s also indicated significance, $t(57) = -2.59$, $p = .012$, but the significance was attributed to just one student. Regression data provided similar findings for math; 36% of the variance could be explained by the combination, but when adjusted for sample size, the variance dropped to 0%, $R^2 = .36$, $\Delta R^2 = .0$, $F(37, 57) = .86$, $p = .69$. The teacher combination 2o, 3s also suggested that further investigation into instructional practice might be appropriate.

School F summary data revealed no significant difference among teacher combinations. The model predicted that 29% of the variance in student achievement could be explained by the combination of teachers in ELA Grades 2 and 3. When adjusted for sample size and number of predictors, the variance explained dropped to 0%, $R^2 = .29$, $\Delta R^2 = 0$, $F(19, 46) = 1.01$, $p = .469$. Of the 20 combinations, 35% included five, six, or seven students. Although not significant at $p < .05$, five teacher combinations that included Teacher 3q as the third-grade teacher suggested a closer look might be warranted. Regardless of the second-grade teacher, the mean score of the 27 students who were linked to Teacher 3q was above the class, EVSC, and state averages on the ISTEP+ test. The distribution of students assigned to Teacher 3q met the assumption of randomness identified in Chapter 1. The data showed similar results in math; 27% of the variance could be explained by the second- and third-grade combination, but when adjusted for sample size, the variance explained dropped to 0%, $R^2 = .27$, $\Delta R^2 = .0$, $F(19, 46) = 92$, $p = .57$. Students linked to Teacher 3q in math also showed higher-than-average ISTEP+ achievement scores.

Summary data for School G noted no significance, $R^2 = .08$, $\Delta R^2 = 0$, $F(3, 25) = .75$, $p = .532$. The regression model predicted that 8% of the variance in student achievement could be explained by the combination of teachers in Grade 2 and Grade 3. When adjusted for sample size and number of predictors, the variance explained dropped to 0%. Because School G was smaller and had fewer sections at each grade, the student count for each combination ranged from five to 10. Two combinations included seven students each. The data for math provided by the regression indicated that 1% of the variance could be explained by the second- and third-grade teacher combination. When adjusted for sample size, the variance was 0% explained, $R^2 = .01$, $\Delta R^2 = .0$, $F(3, 25) = .09$, $p = .97$.

The regression analysis provided statistically significant data for the sample schools in terms of the impact of second- and third-grade teacher combinations on third-grade ISTEP+ scores. Because of the small sample size available, a generalization of significance to the larger population could not be made. As with the ANOVA, sample significance derived from the regression might have the potential to inform school and district policy and teacher placement decisions. Suggestions for further study are included in Chapter 5.

4. *Are third-grade ISTEP+ scores predictable based on first-, second-, or third-grade teacher?* Question 4 was addressed using a simultaneous multiple regression model. Dichotomous-coded variables Y or N identified a student's exposure to a particular teacher. The intention of the fourth question was to explore the possibility that third-grade ISTEP+ scores could be predicted based upon a student's first-, second-, or third-grade teacher. A regression analysis provided summary data to examine.

The summary data for School A indicated that the overall fit of the regression model gave adequate description to the pattern of the data in Grades 1, 2, and 3. R^2 and the adjusted R^2

varied little within each grade level and suggested that the sample size (*n* = 38) was sufficient to produce meaningful data for analysis of ELA and math scores. The regression data revealed that using a .05 level of statistical significance, third-grade ISTEP+ scores could not be predicted based on first-, second-, or third-grade teachers in ELA. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of 85% could be explained by factors other than the grade-level teachers. Similarly, when examining the effect of second-grade teachers as predictors of ISTEP+ scores, a 25% variance was attributed to random unsystematic factors. Finally, when examining the effect of third-grade teachers as predictors of ISTEP+ scores, there may be a collection of variables other than the third-grade teachers that explain 95% of variance among scores.

At .05 level of statistical significance, School A's third-grade ISTEP+ scores could not be predicted based on first-, second-, or third-grade teachers in math. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of 91% could be explained by factors other than the first-grade-level teachers. Similarly, when examining the effect of second-grade teachers as predictors of ISTEP+ scores, a 97% variance was attributed to random factors. Finally, when examining the effect of third-grade teachers as predictors of ISTEP+ scores, the data revealed that variables other than the third-grade teachers explained 29% of variance among math scores.

The R^2 and the adjusted R^2 for School B supported the assumption that the model provided meaningful data for analysis. The F statistic for Grade 1 indicated that there was a statistically significant impact on third-grade ISTEP+ scores based on the teachers in that grade, $F(2, 23) = 3.74$, $p = .039$ in ELA. The regression data revealed that third-grade ISTEP+ scores could be predicted based on second- or third-grade teachers. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of only 3.9% could be explained by factors other than the grade-level teachers. This finding met a .05 level of significance. When examining the effect of second-grade teachers as predictors of ISTEP+ scores, however, a 66% variance was attributed to random unsystematic factors. Finally, when determining the effect of third-grade teachers as predictors of ISTEP+ scores, 51% could not be attributed to third-grade teachers. Hyphenate later paragraphs accordingly.

Third-grade math ISTEP+ scores for School B could not be predicted based on first-, second-, or third-grade teachers at .05 level of statistical significance. When analyzing first-grade teachers as predictors of ISTEP+ scores, a variance of 25% could be explained by factors other than the first-grade level teachers. When examining the effect of second-grade teachers as predictors of ISTEP+ scores, a 97% variance was attributed to random factors. Finally, when examining the effect of third-grade teachers as predictors of ISTEP+ scores, the data revealed that 96% of the variance could be attributed to variables other than the third-grade teachers.

In reviewing the data for School C in ELA, the R^2 and the adjusted R^2 indicated no reason to believe the data were not meaningful, however, the F statistic in each grade revealed no significance in terms of teacher influence on ISTEP+ scores. The regression data indicated that third-grade ISTEP+ scores could not be predicted based on first-, second-, or third-grade teachers. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of 83% could be explained by factors other than the grade-level teachers. Similarly, when examining the effect of second-grade teachers as predictors of ISTEP+ scores, a 73% variance was attributed to random unsystematic factors. When examining the effect of third-grade teachers, 8.6% of the variance could not be attributed to the teachers, $F(2, 30) = 2.665$, $p = .08$. Although notable, the p statistic did not meet the .05 level for significance.

The math data for School C showed a similar pattern. A variance of 35% could be explained by factors other than the grade-level teachers as predictors of ISTEP+ scores at first grade. Similarly, when examining the effect of second-grade teachers as predictors of ISTEP+ scores, a 75% variance was attributed to factors other than the teachers. The effect of third-grade teachers reflected a similar significance that was found in the ELA data. Only 5.3% of the variance could not be attributed to the teachers, $F(2, 30) = 3.243$, $p = .053$. Again, the p statistic did not meet the .05 level for significance, but did suggest that further investigation of the third-grade configuration would be appropriate.

The summary data for School D also appeared to provide adequate information for analysis. The F statistic calculated in ELA for Grades 1 and 2 showed no significance. The statistic for Grade 3, however, indicated substantial impact from third-grade teachers on ISTEP+ scores, $F(4, 58) = 6.248$, $p = .000$. Earlier examination of the ANOVA results suggested third-grade teachers could warrant additional study. The data derived from the regression analysis further supported that inference.

The data provided by the math regression analysis suggested significance in both first grade, $F(5, 57) = 2.42$, $p = .047$ and third grade, $F(5, 57) = 2.58$, $p = .047$. In each of the grades, only 4.7% of the variance of ISTEP+ math scores could be explained by factors other than the grade-level teachers. Second grade showed no predictive significance.

Summary data for School E revealed no significant difference for teachers in Grades 1 and 3. The regression data revealed that third-grade ISTEP+ scores could not be predicted based on first- or third-grade teachers. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of 44% could be explained by factors other than the grade-level teachers and 86% of the variance could not be attributed to third-grade teachers. The *F* statistic denoted significance in second grade, $F(6, 88) = 2.457$, $p = .030$.

Summary data for math revealed no significant difference for teachers in Grade 2, $F(6, 88) = 1.22$, $p = .31$ and Grade 3, $F(5, 89) = .88$, $p = .49$. The regression data, however, did indicate predictive significance in grade one $F(6, 88) = 2.21$, $p = .049$. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of only 4.9% could be explained by factors other than the grade-level teachers.

Summary data for School F indicated no significant difference for teachers in Grade 1 and 2. The regression data revealed that third-grade ISTEP+ scores could not be predicted based on first- or second-grade teachers. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of 13% could be explained by factors other than the grade-level teachers. Similarly, when examining the effect of second-grade teachers as predictors of ISTEP+ scores, an 86% variance was attributed to random unsystematic factors. Finally, when examining the effect of third-grade teachers as predictors of ISTEP scores, only a 1% variance could not be attributed to third-grade teachers. This indicated significance at the .05 level. The *F* statistic did suggest that further examination of Grade 3 data were warranted $F(3, 62) = 4.053$, $p = .011$.

The summary math data for School F revealed no significant predictive difference for teachers in Grades 1 and 2 on ISTEP+ scores. Factors other than the first-grade teachers attributed 65% of the variance. In second grade, 70% of the variance could be explained by random unsystematic factors. When examining the effect of third grade teachers as predictors of ISTEP scores, a variance of only 3.1% could be explained by factors other than the grade-level teachers, $F(3, 62) = 3.15$, $p = .03$. This indicated significance at the .05 level.

The regression summary data for School G suggested that there was no statistical significance among the three grade levels that provided support for predicting ISTEP+ third-grade scores. The regression data revealed that third-grade ISTEP+ scores could not be predicted based on first-, second-, or third-grade teachers. When examining the effect of first-grade teachers as predictors of ISTEP+ scores, a variance of 6.2% could be explained by factors other than the grade-level teachers. The small percentage might merit further investigation. When analyzing the effect of second-grade teachers as predictors of ISTEP+ scores, the percentage was much higher. A 56% variance was attributed to random unsystematic factors. Finally, when examining the effect of third-grade teachers as predictors of ISTEP+ scores, 67% could not be attributed to third-grade teachers.

Data made available by the math regression reflected similar findings. First grade appeared to be more predictive than either second- or third-grade, however, not at a statistically significant level, $F(1, 27) = 2.66$, $p = .114$. A 64% variance in second-grade scores could be explained by random factors not related to the teachers. In third grade, 74% of the variance could be explained by unsystematic factors. ELA and math data are represented in Table 15.

Table 15

Percent of Variance Explained by First-, Second-, and Third-Grade Teachers By School

	School	*N*	Grade 1	Grade 2	Grade 3
ELA	A	38	.849	.244	.950
	B	26	.039*	.666	.509
	C	33	.828	.729	.086
	D	63	.204	.777	.000**
	E	95	.438	.030*	.392
	F	66	.130	.864	.011*
	G	29	.062	.565	.668
Math	A	38	.967	.285	.909
	B	26	.245	.966	.961
	C	33	.345	.750	.053
	D	63	.047*	.958	.047*
	E	95	.049*	.306	.493
	F	66	.653	.699	.031*
	G	29	.114	.643	.744

Note. *$p < .05$; ** $p < .001$

The four types of statistical analysis included in this study provided a multi-factored picture of the measurable impact that first-, second-, and third-grade teachers had on third-grade ISTEP+ scores in a sample of seven elementary schools including 339 students. Statistical results and summary data were provided for each hypothesis that was tested. A discussion of the findings, implications, and practical application of the results are provided in Chapter 5.

CHAPTER 5

SUMMARY OF THE STUDY

This chapter provides a brief summary of the study. Also included is a discussion of the findings and how these findings relate to prior and ongoing research. An examination of the implications of using high-stakes test scores as measures of teacher performance is presented along with suggested possible directions for future studies.

Present student achievement and school accountability models as a typical practice do not take into account students' academic instruction and experiences prior to the initial testing administration in third grade. The push for tying student academic success to teacher performance is a reality, and adequate ways to measure that link is of major concern throughout the country. This study's primary purpose was to add to the body of research that focuses on the student–teacher link and to explore and quantify the contributions to student achievement from previous teachers using scale scores obtained from the third-grade Indiana standardized test. A non-experimental method for estimating teacher effect on student achievement was tested by analyzing the existing scores of a narrow cohort of students. The following criteria offered a basis on which to examine the variance among subgroups and the whole: overall mean achievement scores in both math and ELA separately, as school and grade-level subgroups, and subgroups within the grade level. Linking student Grade 3 standardized assessment outcomes to primary-grade teachers provided empirical data points that offered insight into the practice of evaluating teacher quality and effectiveness of instruction. The summary of the results is discussed in the subsequent sections of Chapter 5.

Limitations

The fundamental challenge to measuring the impact of school factors, specifically teacher impact, is separating their contributions from the many other factors that influence student achievement. Student factors such as intrinsic motivation, home environment, unmeasured prior experiences, and parent engagement impact student achievement gains. A teacher's classroom may reflect a variety of additional factors: principal's bias, preferential treatment of a colleague, ability grouping, and parent requests. The difficulty of measuring and assigning amounts of learning to specific teachers while controlling for these factors is problematic (Rothstein, 2010). High student achievement in a school that serves largely affluent families may give the impression that the teachers within that school are more effective than those teachers in schools that serve children from family backgrounds that are disadvantaged. In order to control for this reality, the study was designed to consider teacher impact on student achievement within a particular school and to avoid school-to-school comparisons, thus mitigating issues of demographic disparity. Students were randomly assigned to teachers who were experienced at that grade level, knew the curriculum, were afforded the same professional development, and had access to the same formative assessment data to assist them in developing lessons that would meet the needs of students in their classrooms. Although significant effort was taken to minimize random unsystematic factors, it was assumed that there continued to be unmeasured factors that influenced academic performance.

Theoretical Implications

It is clear and supported by considerable research that teachers have a profound impact on student achievement (Darling-Hammond, 1997; DuFour, 2004; Dweck, 2006; Hattie, 2008; Marzano, 2003). The question is if a measure of that impact is meaningful and can be generalized among teachers, subject areas, and grade levels.

In this age of accountability, a reliance on the business model has emerged as the vehicle by which to measure teacher performance. Econometric design assigns numerical values to measure a desired result. Throughout the country in the case of academic achievement, high stakes achievement tests administered to students in specific grade levels gauge the acquisition of predetermined knowledge, skills, and processes that students are expected to learn at specific intervals in their educational progression. The concept is that if a student receives adequate instruction and content, he or she will achieve at an expected rate based on his or her age. Simply stated, for every day of classroom instruction by teacher X, there is a quantifiable amount of learning by student Y. In accordance with the business model, legislatures have embraced educational policy that has shifted to a more aggressive focus on the quality of instruction, especially after the signing of the NCLB law in 2002. The business theory runs counter to the generally held notion that a student's environmental circumstance has a far greater impact on achievement than any school-based factors (Coleman et al., 1966). In theory, the business model appears logical and works quite well. In reality, authentic learning is a multifaceted human collaboration (Deci & Ryan, 1985; Dweck, 2000; Marzano, 2003; Wiggins, 2010). Within the learning paradigm, all participants bring with them a confluence of cognition, emotion, past experiences, and innate tendencies.

In an attempt to take into consideration the complexity of classroom dynamics, Indiana has identified achievement metrics in two ways: academic growth and academic achievement. The growth model assigns an amount of expected academic acquisition to each child's yearly progress. The theory is that using this model mitigates or at least takes into account demographic considerations. Students who meet that expectation are designated as having typical academic growth. Those above and below that expectation are assigned high growth or low growth, respectively. The achievement model measures student academic acquisition to a level of mastery designated by the state. The cut score is set each year based on a variety of statistical calculations and assumptions. For the purposes of this study, achievement level scores were the metrics used as the basis of analysis.

Methodological Implications

The study utilized three types of inferential statistical analysis: descriptive, ANOVA, and multiple regression. By using mean achievement scores in ELA and math, the descriptives offered a picture of each school's sample in context to the ISTEP+ scale scores and demographics of EVSC and Indiana as a whole. No attempt was made to compare the cohorts of sample schools within the EVSC.

An ANOVA provided internal comparisons within a school's grade level to detect differences in individual teacher impact on student success not attributed to chance occurrences. Finally, multiple regression analysis addressed the final two questions of the study. Patterns of success that could be attributed to specific teacher combinations were examined and predictability of achievement on ISTEP+ was investigated based on first-, second-, or third-grade teachers.

Relation to Prior Research

In the recent past, measuring teacher impact was directed to Grades 3-12 and as cited in Chapter 2, few studies had investigated the contributions of the teachers prior to the tested grades. The difficulty of including the contributions of teachers who do not teach tested course work continues to be problematic and a variety of models to accommodate that dilemma have been implemented with mixed success. From research in the social sciences, we know that the fine and practical arts, physical education, and engagement in school activities impact student success in school. Linking students' attitudes and mindsets about their abilities to increase academic success are well documented (Deci & Ryan, 1985; Dweck, 2000, 2006).

Summary of the Findings

The results of this study primarily indicated that small sample size created a significant limitation in the study's design. Restrictive parameters were implemented in the attempt to control for random unsystematic variables that impact student academic performance as discussed in the Chapter 2's review of the literature. The typical configuration of elementary Grades 1 through 3 also diminished the sample size. In Indiana, a class size of 18 students is considered optimum in first grade. In Grades 2 and 3, school administrators attempt to keep class size to no more than 27 and a teacher assistant is added at that number; therefore, principals prefer to keep class sizes as small as possible. Filtering out students who were transient, switched teachers during the year, or did not attend the same school for the three years of the study further reduced the number of students who could take part in the cohort.

Even though the descriptive data showed a large number of $z \pm .40$, except in a small number of cases the data for Question 1 revealed that there were no patterns of success that were statistically significant detected in first-, second-, and third-grade teacher combinations reflected in third-grade ISTEP+ scores. There was reason to suspect a variable associated with School D's first grade impacted the scores in four of the six classrooms in a pattern different than randomness. The mean scores in School C's second grade suggested a pattern in math achievement. School F's first-grade math also indicated there may be a pattern. Although providing the reader with statistical descriptions of each grade level within a school, the small sample sizes interfered with identifying patterns of success within grade levels and among teacher combinations. One explanation might be that a single student score could be highly influential and create the appearance of a large differential in mean student achievement data. It is likely that using the mean as the comparative statistic could be problematic.

Notable studies do use the mean standard deviations associated with student achievement to describe teacher effectiveness. If students are randomly distributed into classrooms and the resulting ISTEP+ scores are normally distributed, standard deviation of the scores suggests the magnitude of the teacher effects (Nye et al., 2004). According to Nye et al. (2004), if teacher effects are normally distributed, the difference between the achievement gains between a student having a *not so effective* teacher and an *effective* teacher is over 0.35 of a standard deviation in reading and 0.48 of a standard deviation in math. They contended that the effects are large enough to elicit policy significance. In reviewing the standard deviations found in ISTEP+ descriptive data generated by this study, statistical significance that reflects the magnitude found in Nye et al.'s data was revealed as denoted by asterisks (*) in the Chapter 4 tables.

The ANOVA used in Question 2 provided an additional set of data for interpretation to identify the significance of the teacher's impact on student achievement as measured by ISTEP+. Variance was determined among the teachers in the grade level. The ANOVA data connected the variance to specific teachers in specific grade levels. As illustrated in the Figure 1, student achievement differentials could be easily identified for further analysis. For example, Teachers 3n and 3r were linked to student scores that were considerably different than Teachers 3o, 3p, and 3q. A building administrator would have reason to suspect that something different was occurring within the third-grade classrooms and among the third-grade teachers especially if there was a random distribution of students to classrooms. The *F* statistic provided a data point that would suggest significance; however, the ANOVA supplied the visual to determine whether the difference was linked to positive student achievement data or negative data.

```
Teacher    N    Mean      --------+---------+---------+---------+-
3n         9    420.78    (-------*--------)
3o        16    473.13                      (-----*-----)
3p        16    474.06                      (-----*-----)
3q        13    483.38                       (------*------)
3r         9    430.78       (--------*-------)
                          --------+---------+---------+---------+-
                                 420       450       480       510
```

Figure 1. School D third-grade teachers linked to third-grade ISTEP+ ELA.

There were no significant differences among teachers in School A in either ELA or math indicated in the ANOVA. School B noted significance in first-grade ELA only. No significance was noted in School C, but in School D, first-grade mean scores differed significantly in math. In third grade, scores differed in both ELA and math. Math and ELA scores in School E showed significance in first grade, and in School F, both third-grade subjects were significant at the .05 level. School G's data noted no significance in the analysis in either subject or in grade levels that were part of the study. ANOVA charts are provided in the Appendix to illustrate the significant data points.

The results of Question 3 were most notably affected by the small sample size available for the regression analysis. Because of this limitation, the analysis was narrowed to examine only the second- and third-grade teacher combinations. The data garnered by the regression for the contracted analysis was viewed cautiously. Patterns of success were identified in a limited number of teacher combinations; however, the data did not support the ability to generalize the results to a larger population. Only School E showed significance in ELA at the .05 level for two teacher combinations. In math, School C revealed one combination and School F noted two combinations of significance.

The data analysis for Question 4 focused on the ability to predict student achievement based on grade-level teachers. The analysis produced interesting findings but because of sample size the results were interpreted with restraint. The R and R^2 for each of the sample schools indicated that the sample was large enough to be meaningful for the purposes of the study. The results of the regression presented four points of significance for ELA: School B first grade, School E second grade, School D third grade, and School F third grade. School G was very close to significance in first-grade ELA with $p = .062$. The results for math presented four points of significance as well: Schools D and E first grade and Schools D and F third grade. School C narrowly missed statistical significance with $p = .053$ in third grade. Generally, the cohort analysis provided some evidence that first grade and third grade might have a larger effect on student academic achievement than did second grade.

Among the four analyses, there was no evidence that any individual analysis could stand alone as a measure of teacher effectiveness and student achievement scores. Together, however, a general observation was made. A link could be inferred between first grade or third grade and student success on ISTEP+ in math and ELA. A summary of the descriptives indicated that 52% and 60% of the statistical significance was found in first grade and third grade, respectively. Only 45% was attributed to second grade. The ANOVA indicated that 0% of statistical significance was discovered in second-grade metrics and 57% was detected in either first grade or third grade. The interpretation of Question 3's regression data associated with second- and third-grade teacher combinations also supported the notion that third grade might be the more influential in the pair. Of the significant combinations, the third-grade teacher was the reoccurring factor. Finally, Question 4's data noted three points of significance in Grade 1, four points of significance in Grade 3, and only one point of significance in Grade 2 as represented in Table 15.

Practical Implications

The purpose of the study was to explore the impact first-, second-, and third-grade teachers had on Grade 3 ISTEP+ scores. Although the study did not provide a result that is irrefutable, it does offer implications for practical use at the district and school levels in policy and resource alignment. As noted in the statistical discussions of Questions 1-3, specific teachers surfaced as having positive impact on student achievement. School D Teacher 3s, School E Teacher 1n, and School F Teacher 3q all emerged as foci that require further attention.

Together with formative and summative assessment tools, the administrative leadership responsible for monitoring and supporting student academic achievement can utilize the insight presented in these pages to consider their decisions through an additional lens. Careful consideration of the unintended consequences resulting from using high-stakes test data to measure teacher effectiveness, however, would be wise (Berliner, 2010). Teacher performance based on state test outcomes does not necessarily identify the most effective teachers if the assessment is gauging student conceptual understanding (Rothstein, 2010). Another concern confronting decision makers may be that facing accountability for job performance, teachers would emphasize topics and skills simply to raise test scores (Campbell, 1976). The intense accountability has the capacity to reach all individuals associated with the teaching and learning paradigm. The concerns stated above are increasingly relevant since most states are adopting the Common Core State Standards that assess higher-level conceptual understanding but continue to administer the high-stakes assessment by current standardized tests.

Although the research supports teachers as being the most important factor in student achievement at the school level, the methods by which to measure the ingredients associated with effectiveness continue to be suspect. It is conceivable that in emphasizing test scores there is the chance that districts could reward teachers who do not deserve it and punish those who do (Rothstein, 2011). There is also a real temptation to use formative assessment as performance indicators of teacher effectiveness. Decisions made by teachers, students, and administrators would likely focus on the indicators that achieved the short-term results on the state tests and not necessarily the knowledge and processes that would be of benefit for the student in the long term.

Typically during the early spring, building-level administrators make plans regarding their teaching staff. They assign teachers to specific grade levels or content areas for the upcoming school year. These decisions are based primarily on parameters provided by the collective bargaining agreement, the state licensing mechanism, teacher requests, and principal observations. The findings of the study offer the principals an additional opportunity to focus their thinking more intentionally on what happens in first grade and third grade and specifically which teachers they assign to those classrooms. At the district level, those who direct resources linked to professional development and targeted support efforts may benefit from the study's insight as well.

The study does not identify the characteristics of effective teachers or quantify the attributes associated with teacher quality. A deep understanding of the attributes of effective teaching cannot be overemphasized as an essential element of effective leadership and must be acquired by the building principal. Along with a thorough understanding, the possibility that first grade and third grade emerge as valid indicators of achievement must be considered. Because the study did reveal slight patterns of achievement associated with Grades 1 and 3, additional attention by the administrator in determining teaching assignments might be justified.

Recommendations

The fundamental challenge to measuring the impact of first-, second-, and third-grade teachers on student success continues to be separating contributions from the many factors that influence achievement. The study, although limited, suggested that further explorations that focus specifically on first grade and third grade would be prudent. An analysis of the exiting data on a broader level with fewer limitations placed on characteristics of the study cohort may provide comparative data to determine if the results of this study would generalize to a larger population. Two reasonable follow-up questions generated from the findings of the study could be based on the content or specifically what the students are expected to learn within specific grade levels. Are the first- and third-grade curricula mapped more effectively taking into account the developmental needs of the students and the expectations of the third-grade test than the curriculum in the second grade? Is the difference predicated on a school-level factor such as the Guaranteed and Viable Curriculum (Marzano, 2003)?

One of the most troublesome implications for linking teachers to students in order to measure impact on achievement emerges as a result of the impending statutory requirement that a teacher's evaluation will be based partially on student achievement scores from the state's high-stakes ISTEP+ test. The design of this study intentionally controlled for many of the student and teacher school-level variables by selecting schools that had very stable student populations and teaching assignments. Issues related to transience were addressed so that ISTEP+ scores used were connected to students whose teacher links remained stable during the study. The study also controlled for factors associated with the teacher's competency at teaching a specific set of content expectations at grade level and developmental stages of student learning. Teachers taught the same grade level and curriculum throughout the study. As indicated earlier in this chapter, the sample pool was severely narrowed based on these limitations and could not be generalized to the larger district population.

Changes in content and pedagogy that are based on a substantial collection of research on teaching and learning emphasized the need and expectation that curriculum and instruction be differentiated to meet the needs of each student (Thomlinson, 1999; Thomlinson & Imbeau, 2010). Nationwide, an emphasis on the shift from whole-group to small-group instruction using leveled text is expected, and significant dollars are allocated for professional development that is intended to move teachers toward this goal. Meeting the needs inherent in differentiated teaching requires a school organization that is flexible and has the human capacity to use formative assessment to group and regroup students based on data throughout the year. EVSC and most Indiana school districts do not enjoy that capacity. Even if school districts had the needed teaching staffs, they do not have the electronic capacity to track the ever-changing links between student exposures to specific teachers.

Throughout the state of Indiana, the DIBELS formative assessment is administered four times a year to each kindergarten, first-grade, and second-grade student. Based on the data extracted from the assessment, teachers are to create small groups that provide systematic and targeted differentiated instruction to meet each student's needs within the classroom (Anderson, 1979; Barr et al., 1991; Thomlinson, 1999). Because of capacity issues, EVSC teachers in the primary grades often create these groupings among student in the entire grade level and not just within a teacher's classroom. Often teacher assistants are utilized to provide support with classroom management and also as partners to practice skills with students. Teacher assistants are paid at minimum wage, and there is no standard training in effective instructional strategies. If students are linked to the classroom teacher and then receive targeted instruction from another teacher or support from an assistant to bolster their skills, the link between student achievement and teacher instruction is corrupted immediately.

Further compounding the quandary is the push toward the response to intervention (RTI) model (Fuchs & Fuchs, 2006). EVSC is implementing RTI district-wide. The initial phase was to shore up Tier I, which is the standard curriculum for all students at a grade level. The second phase was to implement Tier II interventions. Tier II interventions are to be available to 20% of the most academically in need students. They also require that additional time be added to a student's instruction based on formative assessment (Davis, Lindo, & Compton, 2007; McCook, 2006). Moreover, a different curriculum and consequently different instruction are components of Tier II. The student groupings are small and research suggests that only four to six students work together with the teacher at any one sitting. Again the teacher-student link that the state intends to use for evaluating teacher performance is indistinct. Tracking student and teacher links are almost impossible if research-based instructional design is implemented with fidelity.

Because the study yielded points at which to begin a more targeted assessment of the teacher impact on student achievement, a relatively simple next step of analysis would be to replicate the study using ISTEP+ Grade 3 scores for 2011-2012. At the very least, discussion regarding the findings of the study in terms of teacher placement in grade-level assignments is warranted.

Another possible next step to broaden the available data for analysis would be to include teachers and students who loop from one grade to the next. Typically the first-grade teacher would loop with the same students to second grade. Even though the practice of looping was not widespread in the EVSC, there were several occurrences that were excluded from the original cohorts and could be examined in a subsequent analysis.

The ultimate goal of educational research is to understand and thus act appropriately to support the multifaceted relationship between teaching and learning. No study or set of data is perfect but they do provide more information that can be used to consider personnel, curriculum, resource, and professional development decisions. This study suggests that factors linked to first grade and third grade might be a logical point at which to begin.

Opportunities for Further Study

The data from the initial study was intriguing and timely in terms of the debate about accountability and teacher performance. An opportunity to delve deeper into the first, second, and third grades and examine their impact on ISTEP+ scores could be achieved through a longitudinal study of the same seven schools and the same cohort of teachers. The longitudinal model would provide an increased *N*. Small sample size in the initial study appeared to be a significant limitation when interpreting the data, especially for Question 3. Using ISTEP+ third-grade scores for the years 2011-2012 and 2012-2013 could provide a solution for that problem.

Assuming random distribution and controlling for student and teacher transience, an analysis of three years of ISTEP+ data might determine with statistical significance if those scores are predictable based on first-, second-, or third-grade teachers. The same model could be used with data from any state as long as the conditions are met. In the future a cross-state comparison would be possible and, with implementation of Common Core State Standards, the results might have national implications. A practical application of the study's results might be the development of a model that would allow schools and/or districts to insert their own data and determine patterns of achievement among grade levels and their teaching staff.

Although the initial study was predicated on a consistent and stable set of students and teachers, a comparison study between the population as a whole and the study's sample would offer insight as well. Questions to be answered could include issues related to student and teacher transience. Implications might be drawn about the teacher's understanding of children and the curriculum at a particular stage of academic development. Student resilience and ability to rebound from ineffective teachers could be examined and possibly determined.

The initial study uncovered more questions than answers, but with the focus on linking student achievement to teacher performance, the consequences of doing so without significant and thorough study are irresponsible. Continuous investigation and scrutiny of standardized tests and their correlation to the processes and models of performance assessment will only ensure that as we learn more our metrics will evolve to be reliable and always valid.

REFERENCES

Aaronson, D., Barrow, L., & Sanders, W. (2007). Teachers and student achievement in the Chicago public schools. *Journal of Labor Economics, 25*(1), 95-135. Retrieved from http://www.chicagofed.org/digital_assets/publications/working_papers/2002/wp2002-28.pdf

Alliance for Excellent Education. (2010). *There is a crisis in America's high schools*. Retrieved from http://www.all4ed.org/about_the_crisis/schools/dropout

American Psychological Association. (2011). *Facing the school dropout dilemma*. Retrieved from http://www.apa.org/pi/families/resources/school-dropout-prevention.aspx

Anderson, L. M. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79*, 193-223. Retrieved from ERIC database. (EJ201788)

Angus, D. L., & Mirel, J. E. (1999). *The failed promise of American high school, 1890-1995.* New York, NY: Teachers College Press.

Bacon, F. (1605). *The advancement of learning*. Retrieved from http://www.iep.utm.edu/bacon/

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs, NJ: Prentice Hall.

Barr, R., Kamil, M., & Mosenthal, P. (1991). *Handbook of reading research* (Vol. 2). Mahwah, NJ: Lawrence Erlbaum Associates.

Barth, P. H. (1999). *Dispelling the myth: High poverty schools exceeding expectation.* Washington, DC: The Education Trust.

Bell, T. H. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office.

Berliner, D. (2010). High-stakes testing and twenty-first century skills. In R. Marzano, *On excellence in teaching* (pp. 112-143). Bloomington, IN: Solution Tree Press.

Berry, B. F. (2007). *Linking teacher and student data to improve teacher and teaching quality*. Retrieved from http://www.dataqualitycampaign.org

Blackwell, L. T. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246-263. Retrieved from ERIC database. (EJ754583)

Boocock, S. S. (1972). *An introduction to the sociology of learning.* Boston, MA: Houghton-Mifflin.

Bornholt, L. J. (1991). *Perceptions of achievement: Compared to whom?* Paper presented to Australian Association for Research in Education, Sydney, Australia. Retrieved from http://www.aare.edu.au/91pap/born191026.txt

Bressoux, P., & Bianco, M. (2004). Long-term teacher effects on pupils' learning gains. *Oxford Review of Education, 30,* 327-345. Retrieved from ERIC database. (EJ681190)

Bronson, P., & Merryman, A. (2010). *The creativity crisis.* Retrieved from http://www.thedailybeast.com/newsweek/2010/07/10/the-creativity-crisis.html

Bruner, J. S. (1960). *The process of education.* New York, NY: Random House.

Bush, J. (2010). *Foundation for excellence in education.* Retrieved from http://wwwexcelined.org

Campbell, D. T. (1976). *Assessing the impact of planned social change.* (Paper #8) Hanover, NH: The Public Affairs Center, Dartmouth College. Retrieved from https://www.globalhivmeinfo.org/CapacityBuilding/Occasional%20Papers/08%20Assessing%20the%20Impact%20of%20Planned%20Social%20Change.pdf

Carnegie Commission. (1991). *In the national interest: The federal government in the reform of K-12 math and science education.* New York, NY: Author.

Center for American Progress. (2009). *Stimulating excellence.* Retrieved from http://www.americanprogress.org/issues/2009/5/pdf/education_entrepreneurs.pdf

Center on Teaching and Learning. (n.d.). *DIBELS data system.* Eugene, OR: Author. Retrieved from http://ctl.oregon.edu

Charters, W. W., Jr. (1963). The social background of teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 270-294). Chicago, IL: Rand McNally.

Chetty, R., Friedman, J., & Rockoff, J. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood.* (Working Paper). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://nber.org/papers/w17699

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools.* Washington, DC: The Brookings Institute.

Cimpian, A., Arce, H., Markham, E., & Dweck, C. (2007). Subtle linguistic cues affect children's motivation. *Psychology Science, 18,* 314-316.

Cimpian, A. (2010). The impact of generic language about ability on children's achievement motivation. *Developmental Psychology, 46,* 1333-1340.

Coch, D. (2010). Constructing a reading brain. In D. Sousa (Ed.), *Mind, brain, & education: Neuroscience implications for the classroom* (pp. 139-162). Bloomington, IN: Solution Tress Press.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., & Weinfeld, F. D. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Corliss, R. (2010, September 29). Waiting for 'Superman': Are teachers the problem? *Time Entertainment.* Retrieved from http://www.time.com/time/arts/article/0,8599,2021951,00.html

Costa, A., & Kallick, B. (2008). *Habits of the mind.* Alexandria, VA: Association for Supervision and Curriculum Development.

Cotton, K. (1995). *Effective schooling practices: A research synthesis.* Portland, OR: Northwest Regional Educational Laboratory.

Creemers, B. P. (1994). *The effective classroom.* London, England: Cassell.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* New York, NY: Harper Collins.

Danielson, C., & McGreal, T. (2000). *Teacher evaluation to enchance professional practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching.* New York, NY: National Commission on teaching and America's Future.

Data Quality Campaign. (2012). *Data quality campaign executive summary*. Retrieved from http:// www.dataqualitycampaign.org/stateanalysis_executive summary

Davis, G. C., Lindo, E. J., & Compton, D. L. (2007). Children at risk for reading failure: Constructing an early screening measure. *Teaching Exceptional Children, 39*, 32-37.

DeBoer, G. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching, 37*, 582-601. doi: 10.1002/1098-2736

Deci, E., &. Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York, NY: Plenum Press.

Department of education organization act of 1979. Public Law 96-88, 93 Stat. 668, 1979.

DuFour, R. (2004). Schools as learning communities. *Educational Leadership, 61*(8), 6-11.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development.* Philadelphia, PA: Psychology Press.

Dweck, C. S. (2006). *Mindsets: The new psychology of success.* New York, NY: Random House.

Ericsson, K. A. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson (Ed.), *Toward a general theory of expertise: Prospects and limits* (pp. 1-39). Cambridge, England: Cambridge University Press.

Evansville Vanderburgh School Corporation and Evansville Teachers Association. (2010). *Collective bargaining agreement by and with the Evansville Vanderburgh School Corporation and the Evansville Teachers Association. Collective Bargaining Agreement 2011-2014* . Evansville, IN: Author.

Fan, X. T., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review, 13*, 1-22. Retrieved from ERIC database. (ED430048)

Ford, H. (n.d.). *Brainy quote.* Retrieved from http://www.brainyquote.com/quotes/quotes/h/henryford122817.html

Fuchs, D., & Fuchs, L. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99. doi:10.1598/RPO.4141.1.4

Garii, B. (2006). Inside teaching: How classroom life undermines reform. *Issues in Teacher Education, 15*(2), 82-86.

Ghandi, M. (n.d.). *Mahatma Gandhi quotes.* Retrieved from http://thinkexist.com/quotation/if_i_have_the_belief_that_i_can_do_it_i_shall/148499.html

Goe, L., &. Holdheide, L. (2011). *Measuring teachers' contributions to student learning growth for nontested grades and subjects.* Washington, DC: National Comprehensive Center for Teacher Quality.

Good, T. (2010). Forty years of research on teaching 1968-2008: What do we know now that we didn't know then? In R. Marzano (Ed.), *On excellence in teaching* (pp. 30-62). Bloomington, IN: Solution Tree Press.

Gordon, R. K. (2006). *Identifying effective teachers using performance on the job.* Washington, DC: The Brookings Institution.

Grelbrick, J. (1999). *American education*. Retrieved from http://oregonstate.edu/instruct/ed416/ae2.html

Guskey, T. (2003). Using data to improve student achievment. *Educational Leadership, 60*(5), 6-11.

Hampel, R. L. (2008). *Progressive education*. Retrieved from http://www.faqs.org/childhood/Pa-Re/Progressive-Education.html

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84-117.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Association, 100*(2), 1-11. doi:10.1257/aer.100.2.267

Hanushek, E., Jamison, D., Jamison, E., & Woessmann, L. (2008). *Education and economic growth*. Retrieved from http://www.educationnext.org

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes.

Hattie, J. (2003). *Teachers make a difference: What is the research evidence?* Retrieved from http://www.acer.edu.au/enews/2004/05/what-makes-a-good-teacher

Hattie, J. (2008). *Visible learning: a sysnthesis of over 800 meta-analyses relating to achievement.* London, England: Routledge.

Haycock, K. (1998). *Good teaching matters . . .a lot.* Washington, DC: The Education Trust.

Henig, J. (2008). *What do we know about the outcomes of KIPP schools?* East Lansing, MI: The Great Lakes Center for Education Research & Practice. Retrieved from http://www.greatlakescenter.org/docs/Policy_Briefs/Henig_Kipp.pdf

Hess, F. M. (2009). The new stupid. *Educational Leadership, 66*(4), 12-17. Retrieved from ERIC database. (EJ820801)

Hess, F. M., & Fullerton, J. (2010). The numbers we need: How the right metrics could improve K-12 education. *Education Outlook, 2*, 1-8. Retrieved from ERIC database. (ED509600)

Hinton, C., & Fischer, K. W. (2008). Research schools: Grounding research in educational practice. *Mind, Brain, and Education, 2*, 157-160.

Hurd, P. D. (1958). Science literacy: Its meaning for American schools. *Educational Leadership, 16*, 13-16.

Indiana Department of Education. (2011a). *A-F accountability.* Retrieved from http://dow.in.gov/improvement/accountability/f-accountability

Indiana Department of Education. (2011b). *How rise works* [PowerPoint slides]. Retrieved from http://www.riseindiana.org

Indiana Department of Education. (2011c). *ISTEP+ program manual. Policies and procedures for Indiana's assessment system.* Retrieved from http://www.doe.in.gov/sites/default/files/assessment/2011-12-istep-program-manual11-30-11final-kc.pdf.

Indiana Department of Education. (2011d). *Senate Enrolled Act 1.* Retrieved from http://www.in.gove/legislative/bills/2011/SE/SE0001.1.html

Indiana Department of Education. (2011e). *Student learning objectives addendum. RISE: Evaluation Development System*. Indianapolis, IN: Indiana Department of Education.

Indiana Department of Education. (2012a). *Accreditation*. Retrieved from http://www.doe.in.gov/student-services/accreditation

Indiana Department of Education. (2012b). *Data collection and reporting.* Retrieved from http://www.doe.in.gov/improvement/data-reporting

Indiana Department of Education. (2012c). *Office of student assessment*. Retrieved from http://www.doe.in.gov/achievement/assessment

Ingvarson, L., & Rowe, K. (2007). *Conceptualizing and evaluating teacher quality: Substantive and methodological issues.* Retrieved from http://research.acer.edu.au/learning_processes/8

James, A. N. (2007). *Teaching the male brain: How boys think, feel, and learn in school.* Thousand Oaks, CA: Corwin Press.

Jencks, C., Smith, M. S., Ackland, H., Bane, M. J., Cohen, D., Grintlis, H., Heynes, B., & Michelson, S. (1972). *Inequality: A reassessment of the effects of family and schooling in America.* New York, NY: Basic Books.

Jordan, H., & Mendro, R. (1997). *Teacher effects on longitudinal student achievement: A preliminary report on research on teacher effectiveness.* Indianapolis, IN: National Evaluation Institute.

Kamins, M., &. Dweck, C. S. (1999). Person vs. process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology, 35,* 835-847. doi: 10.1037/0012-1649.35.3.835

Kane, T., & Staiger, D. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation.* (Working Paper 14607). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14607

Kashdan, T. (2009). *Discover the missing ingredient to a fullfilling life.* New York, NY: Harper Collins.

Kendall, J. S. (1996). *Content knowledge: A compendium of standards and benchmarks for K-12 education.* Aurora, CO: McRel.

Ladd, H., Noguera, P., & Payzant, T. (2011, November 4). Presidential address to the Association for Public Policy Analysis and Management. Washington, DC: Economic Policy Institute. Retrieved from http://www.boldapproach.org

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, IL: University of Chicago Press.

Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us*. Aurora, CO: Mid-continent Research for Education in Learnng.

Marzano, R. J. (2003). *What works in schools: Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2005). *The art and science of teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2008). *Getting serious about school reform.* Bloomington, IN: Marzano Research Laboratory.

Marzano, R. J. (2010). *On excellence in teaching.* Bloomington, IN: Solution Tree Press.

Marzano, R., McNulty, B., & Waters, T. (2003). *School leadership that works; From research to results.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mayhew, K. C., & Edwards, A. C. (1936). *The laboratory school of the University of Chicago, 1896-1903.* New York, NY: Appleton-Century. Retrieved from http://www.archive.org/stream/deweyschoolthela008095mbp/deweyschoolthela008095mbp_djvu.txt

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: Rand.

McCook, J. E. (2006) *The RtI guide: Developing and implementing a model in your schools.* Horsham, PA: LRP Publications.

Meisels, S. A. B. (2003). Creating a system of accountability: The impact of instructional assessment on elementary children's achievement test scores. *Educational Policy Analysis Archives, 11*(9), 1-18. Retrieved from ERIC database. (EJ680080)

Mendro, R. J. (1998, April). *Longitudinal teacher effects on student achievement and their relations to school and project evaluation.* Presentation at the Annual Meeting of the American Educational Research Association. San Diego, CA.

Moroney, T., & Young, J. (2010, October 13). *Rhee quits as Washington, DC schools chief amid clash with teacher.* Retrieved from http://www.bloomberg.com/news/2010-10-13/michelle-rhee-is-said-to-step-down-as-washington-d-c-schools-chancellor.html

Mueller, C. M. (1998). Intelligence praise can undermine motivation and performance. *Journal of Personality and Social Psychology, 75*, 33-52. doi: 10.1037/0022-3514.75.1.33

National Conference of State Legislatures. (2012). *News room: Issues and research.* Retrieved from http://www.ncsl.org/issues-research/educ/compulsory-education-overview

National Governors Association. (2009). *Common core state standards development work group and feedback group announced.* Retrieved from http://www.nga.org/cms/home/news-room/news-releases/page_2009/col2-content/main-content-list/title_common-core-state-standards-development-work-group-and-feedback-group-announced.html

National Institute for Excellence in Teaching. (1999). *The system for teacher and student advancement.* Retrieved from http://www.tapsystem.org

National Reading Panel. (2000). *Report of the National Reading Panel-teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved from http://www.nichd.nih.gov/publications/nrp

Noe, C. (2004, January 9). *Bush decries democrats' 'soft bigotry of low expectations.'* Retrieved from http://archive.newsmax.com/archives/articles/2004/1/9/110923.shtml

Novak, J., & Cañas, A. (2002). *The theory underlying concept maps and how to construct and use them.* Pensacola, FL: Florida Institute for Human and Machine Cognition.

Nye, B., Konstantopoulous, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26,* 237-257. Retrieved from ERIC database. (EJ727526)

Penfield, R. D. (2010). Test-based grade retention: Does it stand up to professional standards for fair and appropriate test use? *Educational Researcher, 39*(2), 110-119. Retrieved from ERIC database. (EJ886499)

Perrone, V. (1991). On standardized testing. *Childhood Education, 67*(3), 132-142. Retrieved from ERIC database. (EJ430344)

Piaget, J. (1954). *The construction of reality in a child.* New York, NY: Basic Books.

Pink, D. (2005). *A whole new mind.* New York, NY: Riverhead Books.

Pink, D. (2009). *Drive.* New York, NY: Riverhead Books.

Posner, M. I. (2010). Neuroimaging tools and the evolution of educational neuroscience. In D. Sousa (Ed.), *Mind, brain, & education: Neuroscience implications for the classroom* (pp. 27-44). Bloomington, IN: Solution Tree Press.

Ravitch, D. (2010). Dictating to the schools. *Virginia Journal of Education*, 4-9. Retrieved from http://www.veanea.org/home/907.htm

Reeves, D. B. (2006). *The learning leaders.* Alexandria, VA: Association for Supervision and Curriculum Development. doi

Rhee, M. (2010). *Students first: A movement to transform public education.* Retrieved from http://www.studentsfirst.org

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica, 73,* 417-458. Retrieved from http://www.econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf

Robinson, S. K. (2010). *RSA animate: Changing education paradigms*. Retrieved from http://www.youtube.com/watch?v=zDZFcDGpL4U

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York, NY: Holt, Rinehart, & Winston.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics, 25,* 175-214.

Rothstein, J. (2011). *Review of learning about teaching*. Boulder, CO: National Education Policy Center.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12,* 247-256.

Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future students academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research Center.

Sass, T. R. (2008). *The stability of a value-added measures of teacher quality and implications for teacher compensation policy.* Washington, DC: The Urban Institute.

Schlechty, P. C. (2000). *Shaking up the school house: How to support and sustain educational innovation.* San Francisco, CA: Jossey-Bass.

Schlechty, P. C. (2002). *Working on the work.* San Francisco, CA: Jossey-Bass.

Schmoker, M. (2006). *Results now.* Alexandria, VA: Association for Supervision and Curriculum Development.

Schmoker, M. (2011). *Focus: Elevating the essentials to radically improve student learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Socioeconomic status. (2012). In Wikipedia, the free encyclopedia [online.] Retrieved from http://en.wikipedia.org/wiki/Socioeconomic_status

Sousa, D. A. (2010). How science met pedagogy. In D. A. Sousa (Ed.), *Mind, brain, & education: Neuroscience implications for the classroom* (pp. 9-26). Bloomington, IN: Solution Tree Press.

Sprenger, M. (2005). *How to teach so students remember.* Alexandria, VA: Association for Supervision and Curriculum Development.

Sprenger, M. (2010). *Brain-based teaching in the digital age.* Alexandria, VA: Association for Supervision and Curriculum Development.

Thomlinson, C. (1999). *The differentiated classroom.* Alexandria, VA: Association for Supervision and Curriculum Development.

Thomlinson, C. & Imbeau, M. (2010). *Leading and managing a differentiated classroom.* Alexandria, VA: Association for Supervision and Curriculum Development.

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times.* San Francisco, CA: Jossey-Bass.

U.S. Bureau of the Census. (2011). *Historical statistics of the United States: Colonial times to 1970.* Retrieved from http://www.census.gov/prod/www/abs/statab.html

U.S. Department of Education. (2001). *Reading first.* Retrieved from http://www2.ed.gov/programs/readingfirst/legislation.html

U.S. Department of Education. (2009). *Press releases: President Obama, U.S. secretary of education Duncan announce national competition to advance school reform.* Retrieved from http://www2.ed.gov/news/pressreleases/2009/07/07242009.html

U.S. Department of Education. (2010). *Digest of education statistics 2010.* Retrieved from http://nces.ed.gov/pubs2011/2011015.pdf

U.S. Department of Education. (2010). *Turning around the dropout factories.*Retrieved from http://www2.ed.gov/about/overview/budget/budget13/crosscuttingissues/highschoolgrad.pdf

U.S. Department of Education. (2011). *Goals 2000: Educate America act 1994.* Retrieved from http://www2.ed.gov/legislation/GOALS2000/TheAct/intro.html

U.S. Department of Education. (2012a). *Title 1.* Retrieved from http://find.ed.gov/search?q=SES&client=default_frontend&output=xml_no_dtdproxystylesheet=default_frontend&sa.x=22&sa.y=3

U.S. Department of Education. (2012b). *Reading first*. Retrieved from http://www2.ed.gov/programs/readingfirst/index.html

U.S. Department of Veterans Affairs. (2012). *Servicemen's Readjustment Act of 1944.* Retrieved from http://www.gibill.va.gov/benefits/history_timeline/index.html

Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need - and what we can do about it.* New York, NY: Basic Books.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness* (2nd ed.). New York, NY: The New Teacher Project. Retrieved from http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Welch, F. (1974). Relationships between income and schooling. *Review of Research in Education, 2,* 179-201. doi: 10.3102/0091732X002001179

Westbrook, R. (1991). *John Dewey and American democracy.* Ithaca, NY: Cornell University Press.

White, K. R. (1982). The relationship between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461-481. doi: 10.1037/0033-2909.91.3.461

Wiggins, G. (2010). What's my job? Defining the role of the classroom teacher. In R. Marzano (Ed.), *On excellence in teaching* (pp. 7-29). Bloomington, IN: Solution Tree Press.

Wiggins, G. (2011). A true test toward more authentic and equitable assessment. *Kappan Classic, 92*, 81-87.

Willis, J. (2010). The current impact of neuroscience on teaching and learning. In D. Sousa (Ed.), *Mind, brain, & education: Neuroscience umplications for the classroom* (pp. 45-66). Bloomington, IN: Solution Tree Press.

Wong, K. (2004). *District-wide framework for improvement.* Nashville, TN: Vanderbilt University.

Woodhead, M. (2006). *Changing perspectives on early childhood: Theory, research and policy.* Background paper prepared for the Education for All Global Monitoring Report 2007 Strong foundations: early childhood care and education Retrieved from http://unesdoc.unesco.org/images/0014/001474/147499e.pdf

Wright, S. P., Horn, S., & Sanders, W. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67. Retrieved from ERIC database. (EJ548364)