THE IMPACT OF SCHOOL-WIDE PROFESSIONAL DEVELOPMENT ON STUDENT ACHIEVEMENT: A META-ANALYSIS FROM AN EFFECTIVE SCHOOLS PERSPECTIVE

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF EDUCATION

BY

CAMERON L. RAINS

DISSERTATION ADVISOR: DR. JOSEPH MCKINNEY

BALL STATE UNIVERSITY

MUNCIE, INDIANA

JULY 2015

Acknowledgements

Completing this dissertation has been quite a journey and one that I certainly would not have been able to complete without the support of many individuals. While it would be impossible to mention all of them, I would like to acknowledge and thank some of the individuals who helped make it possible for me to endure this work.

First, to my wife Juli, you will always be the patron saint of wives in my opinion. Your patience, encouragement, understanding, and willingness to sacrifice so much in order to help me start and finish this will always serve as the shining example of the selflessness and putting others first attitude that I hope to someday exhibit and possess. I thank my sons, Caden and Jace, for their patience in dealing with a tired dad when late night dissertation work needed to occur and my newly born daughter Kendall for the additional motivation to finish this that only a newborn could provide.

I owe a debt of gratitude to my dissertation chairperson Dr. Joseph McKinney for his support, encouragement, and advice throughout this process. Thank you for helping to steer me in the right direction and ensuring that I finished what I started. In addition, I thank Dr. Serena Solloum for serving as a member of my committee and for her feedback through the multiple iterations of this study as well as her steadfast belief in me. Your detailed feedback, expertise, and constant encouragement were appreciated more than you know.

I would like to extend special thanks to Dr. Robert Marzano for helping conceive the idea of this dissertation and for his mentorship. Your work has had a profound impact on me as an educator and researcher. Your agreeing to serve on my dissertation committee was a bit of a dream come true.

I am appreciative for the feedback and guidance from my committee members, Dr. Jeff Swensson and Dr. Kevin Gerrity. Thank you for pushing my thinking and providing valuable feedback along the way.

A special thanks to Debra Prenkert, who provided encouragement and support throughout the process. Thank you for letting me vent, pick your brain, and celebrate the small victories while on our dissertation journeys. Going through this together kept me motivated. I appreciate your friendship and am inspired by the educator that you are.

Finally, thank you to my parents Randy and Debbi for their lifelong dedication to education and for teaching me that rewards come with hard work. I appreciate your perseverance, as that particular lesson required multiple re-teaching opportunities. In addition, thank you for showing me that learning never ends.

Table of Contents

Tables and Figures

Chapter 1

Introduction to the Study/Background

The primary function of schooling is student learning. For centuries, children have attended schools in the United States with the goal of gaining the knowledge and skills necessary for success in life. Schools, in turn, have a responsibility to ensure that students achieve that goal. Robert Marzano writes, "The ultimate criterion for expert performance in the classroom is student achievement. Anything else misses the point" (Marzano, Frontier, & Livingston, 2011, p. 8). Since the No Child Left Behind Act was passed in 2002, schools all over the United States have been working to meet this criterion and for good reason.

Statement of the Problem

Over the last several decades, there has been considerable increased stress on schools to help students achieve greater academic success. Events like Sputnik, the first satellite in space launched by Russia in 1957, contributed to the dissatisfaction with the quality and performance of public schools. The critical report, "A Nation at Risk" in 1983 was a catalyst for public school reform.

The United States' standing as a global superpower has not been significantly challenged to date economically or militarily, but much worry has been generated by international exams like the Trends in International Mathematics and Science Studies (TIMSS) and the Programme for International Student Assessment (PISA) that show the United States lagging behind in terms of student achievement among industrialized nations (National Center for Education Statistics, 2010; National Center for Education Statistics, 2012).

In the TIMSS Study (2011), the United States ranked below the international average in eighth grade mathematics as well as twelfth grade mathematics and science. There were 57

countries that participated in the fourth grade TIMSS assessment and 56 countries that participated in the TIMSS assessment for eighth graders in 2011. The overall results were as follows:

- In fourth grade mathematics, 7 countries scored higher than the United States while 4 were not measurably different, and 45 scored lower
- In fourth grade science, 3 countries scored higher than the United States, while 6 were not measurably different and, 47 scored lower.
- In eighth grade mathematics, 11 countries scored higher than the United States, while 13 were not measurably different and, 31 scored lower.
- In eighth grade science, 12 countries scored higher than the United States, while 10 were not measurably different and, 33 scored lower (National Center for Education Statistics, 2012).

In 2009, the U.S. PISA results were as follows:

- In reading literacy, 9 countries had higher scores, 16 were not measurably different, and 39 were lower.
- In mathematics literacy, 23 countries had higher scores, 12 were not measurably different, and 9 were lower.
- In science literacy, 18 countries had higher scores, 13 were not measurably different, and 33 were lower (National Center for Education Statistics, 2010).

The failure of the United States to be the top performer has produced a great deal of stress on the education system. The public discourse has moved from schools needing to do the best they can with students toward the expectation that schools produce student achievement gains. Tension between the public, legislators, and schools seems to continue to mount with

each day that passes without significant improvements in student achievement results. The No Child Left Behind Act (2001) further contributed to the demands on schools with the accountability systems that were developed and utilized by each state.

Significance of the Study

There is a substantial research base focused upon increasing student achievement. One of the main barriers for educators is honing in on what we know to be effective and taking these effective strategies to scale at the school and district level. As Edmonds (1979b) points out, we already know more than we need to in order to ensure schools are effective. Taking effective strategies to scale has been very difficult for schools due to the contextual variables that could determine the appropriateness of strategies, the differences in school cultures, values, beliefs, and traditions that affect the effectiveness of strategies, and the vast differences that exist between learners within and between schools. Although there is extensive research on effective schools, to date there has not been a full research synthesis (or meta-analysis). There have been so many studies on effective schools or schools that generate high levels of student achievement, it is sometimes difficult for educators to navigate the research. Knowing how much of an impact or effect a given strategy has on student learning will help practitioners make better decisions about where to focus their efforts. This in turn can lead to taking these strategies to scale more quickly and purposefully.

Purpose of the Study

The purpose of this study was to compile data from effective schools correlate areas or areas that effective schools researchers identified as making a difference for student learning and perform a meta-analysis. Providing a research synthesis will provide educators with a snapshot of much of the research on effective schools in one place. The results from the meta-analysis

will also help educators make decisions about what strategies should be implemented in their context and what kind of results they might expect. There are implications for researchers as well. Due to the sheer volume of effective schools studies, it takes a lot of time to effectively synthesize the information. This research synthesis will provide a good start on this for the research community, allowing future research to be targeted more quickly.

There are two additional benefits to performing a meta-analysis on Effective Schools Research. Combining all of the research from the beginning of the effective schools movement in 1970 through the present generates additional areas of value (what were originally called correlates). It is important for educators to be aware, pay attention to, and plan for improvement in these areas. Also, compiling all of the results and measuring the effect sizes, or the degree to which each strategy influences the normal distribution in the identified categories, elucidates the impact these areas have on student achievement. This will provide the field with the opportunity to consider the broader scope of the effective schools research results in a quantifiable way. In addition, it will show the degree to which staff expectations of student learning and staff professional development influence student achievement. Effect sizes will be generated in the various categories that will help to accomplish this. With this data in hand, educators can make better decisions about which strategies to utilize and how to implement them in ways that may generate student achievement results.

Research Design

To complete this study, meta-analysis, also known as research synthesis was utilized. According to Cooper (2010) a research synthesis, or meta-analysis, summarizes previously done empirical studies and draws conclusions based upon all of them. The goal of a research synthesis is to present the current state of knowledge in the given topic through combining all of

the quantitative data available. This makes meta-analysis both unique and helpful. As noted by Borenstein, Hedges, Higgins, and Rothstein (2009), meta-analysis is much different than traditional narrative reviews of the research in that it weights studies based upon mathematical criteria. Criteria for a meta-analysis are pre-established before the meta-analysis begins, whereas in traditional narrative reviews, the expert researcher typically assigns a level of importance to each study. The use of meta-analysis was ideal for effective schools research because there is such a large research-base on the topic and not a lot of meta-analytic work has been done in the area of effective schools research.

Research Questions

A comprehensive review of the effective schools literature was previously conducted and is discussed as a part of this dissertation in Chapter 2. The meta-analysis for this dissertation was performed based upon data from two of the identified nine variables of effective schools[1]. The nine most recently identified variables were: effective leadership, developing and maintaining a pervasive focus on learning, producing a positive culture, creating high (and appropriate) expectations for all, monitoring progress at all levels, the process of effective teaching, involving parents in productive and appropriate ways, developing staff skills at the school site, and emphasizing student responsibilities and rights. The nine variables are discussed in great depth in Chapter 2. The two areas of focus for this dissertation were "high expectations" and "staff development". For a meta-analysis on two additional areas, "collaboration with a focus on student learning" and "progress monitoring", please refer to Prenkert (2015). The four main research questions for this study based upon effective schools research are:

[1] Information regarding the nine variables is discussed in great depth in Chapter 2.

1. What is the impact that creating high and appropriate expectations for all students has on student achievement based on the literature?
2. What is the impact that developing staff skills at the school site have on student achievement based on the literature?
3. Does the varying duration of staff professional development have an impact on student achievement based on the literature?
4. Of the two variables studied, which has a greater impact on student achievement based on the literature?

Influences of No Child Left Behind

The No Child Left Behind Act (2001) increased emphasis on using research to guide practice in education. The term "scientifically based research" is used 119 times in the act. According to Dahlkemper (2003), The No Child Left Behind Act provided the impetus for schools to find programs that increase student achievement through scientifically based research validated programs. Finding and utilizing school programs that were scientifically research-validated programs represented a new and often very different approach for schools.

With the passage of No Child Left Behind, there was increased pressure on schools to use evidence-based practice to generate measurable student achievement results. Another byproduct of the law was increased accountability for schools with severe penalties and punishment for those that did not achieve, including shutting the school down.

Influences of Race to the Top

In addition to No Child Left Behind, the Race to the Top Grant (2009) further encouraged schools to use evidence-based practice. Originally, in 2010, states had to conform to certain practices in order to be eligible for these federal funds. This continued for the competitive grants

awarded in 2011 and 2012. The Race to the Top funding was originally made available through the American Recovery and Reinvestment Act (ARRA). According to the executive summary, there were $4.35 billion of funds available in the competitive Race to the Top Grant (United States Department of Education, 2009). The goal of the grant was to give states incentives to invest in reform and innovation, improve student outcomes, increase graduation rates, eliminate achievement gaps, and increase college and career readiness. There were four reform areas that were at the center of Race to the Top. They were:

1) Adopting, implementing, and assessing college and career ready standards.
2) Creating data systems that were sensitive to student growth and achievement that could also provide formative feedback to teachers and principals on ways to improve instruction.
3) Hiring, growing, rewarding, and keeping the most effective teachers and building leaders and ensuring they were in place in schools that needed them the most.
4) Identifying and turning around the lowest performing schools.

Race to the Top was designed to accelerate reforms in states across the nation. The hope was that the states initially selected would serve as models for other states so that these reforms would spread across our nation (United States Department of Education, 2009). It is too early to tell if these reforms had any lasting impact on student learning.

While states, and in turn schools certainly had a vested interest in becoming eligible for the federal funds available from the Race to the Top Grant, the massive infusion of money at the federal level increased pressure on schools to implement reform initiatives. This was the case for many schools, even if they did not receive RTTT funds because many states, districts, and schools put policies in place to become eligible. With this kind of focus on the four core

education reform areas one has to question whether these were even the right areas of focus.

With the condition of elevated pressure on schools to increase the use of evidence based practice, and after a thorough review of the research on effective schools, it is clear that it would be beneficial to conduct a research synthesis. Pulling the empirical studies on effective schools, synthesizing the results, identifying key areas that generated positive results on student achievement, and drawing conclusions based on all of the information will be very valuable to the field of education in many ways. This will save educators and policy makers time by compiling results, it will give them direction regarding the most effective building-wide strategies, and it will lay the road map for building effective schools.

For these reasons, I chose to pursue a research synthesis also known as meta-analysis on effective schools for my dissertation. Specifically, I performed a meta-analysis on the available effective schools research. I selected this area because it is an area I am very passionate about, and a research synthesis is needed. One of the biggest questions administrators grapple with in their jobs is: What should be done to increase student achievement and close achievement gaps? A very sizeable research base exists in the area of effective schools. There are at least 9,516 studies that use the terms "effective schools" or "school effects" alone[2].

Definition of Terms

- *Effective School*: a school with high overall student achievement with no significant gaps in that achievement across the major subgroups in the student population (Lezotte & Snyder, 2011).

[2] Data retrieved from ERIC-EBSCO Host on March 23, 2013

- *Effective School Research*: the scientific body of information available regarding effective schools.
- *High Expectations for Students:* The expectancy that all students will learn at high levels regardless of demographic characteristics, previous experiences, or prior learning.
- *Developing Staff Skills:* Ensuring that staff members get better at their craft and is often referred to as professional development.
- *Site Based Staff Development:* a subcategory of developing staff skills, this focuses on staff development at the school level.
- *Ongoing Professional Development:* a subcategory of developing staff skills, the focus is on staff development activities that are ongoing rather than one-shot opportunities.
- *Student Achievement:* Student performance on a standardized measure.
- *Meta-analysis*: A quantitative synthesis of data from multiple studies on a given intervention that combines the information and produces an effect size.
- *Effect Size:* the magnitude of change in relationship to the normal distribution that a given intervention can be expected to produce.
- *School Effects Research:* studies of the scientific properties of school effects evolving from input-output studies to current research utilizing multilevel models (Reynolds et al., p. 3)
- *Effective Schools Research:* research concerned with the process of effective schooling, evolving from case studies of outlier schools through to contemporary studies merging qualitative and quantitative methods in the simultaneous study of classrooms and schools (Reynolds et al., p. 3).

- *School Improvement Research:* examining the processes whereby schools can be changed utilizing increasingly sophisticated models that have gone beyond simple applications of school effectiveness knowledge to sophisticated ‘multiple lever’ models (Reynolds et al., p. 3).

Chapter 2

Introduction

The history of effective schools research began in 1966 and continues today. This chapter highlights the phases of the effective schools movement, identifies many of the most prominent researchers in this arena, discusses how the research was conducted, and illuminates the sheer volume of literature on effective schools. A rationale for a research synthesis or meta-analysis in the area of effective schools research is then elaborated.

Background on Effective Schools Research

Research rapidly increased in the area of effective schools after the release of the momentous study *Equality of Educational Opportunity*, otherwise known as the Coleman Report (Coleman, Campbell, Hobson, McPartland, Mood, & Weinfield, 1966). In short, Coleman and colleagues asserted that family background was the most important factor for success or failure in school, and schools had very little impact. The authors wrote,

> It is known that socioeconomic factors bear a strong relation to academic achievement. When these factors are statistically controlled, however, it appears that differences between schools account for only a small fraction of the differences in pupil achievement (Coleman et al., 1966, pp 21-22).

Coleman's conclusions were corroborated by Jencks and colleagues (1972) in their work, *Inequality: A reassessment of the effect of family and schooling in America.* This work re-analyzed the original data from the Coleman Report. Jencks et al., (1972) found that schools do very little to lessen the gap between rich and poor students, or between more able and less able students. They also found that student achievement is primarily the function of one factor, the background of the student, and that there was little evidence that education reform could improve the influence school has on student achievement.

The conclusions from the Coleman and Jencks Reports created a great deal of angst for school reformers who believed that schools were the best chance to level the playing field and ensure that students were prepared to be successful in life. This sparked great debate in the educational research community. The response was over 47 years of research that strongly disputed the conclusions of the Coleman and Jencks Reports.

Phases and Strands of the Effective Schools Movement

There were a few prominent researchers who set out to challenge the conclusions drawn by the original Coleman study. Among the early researchers were Ron Edmonds, Lawrence Lezotte, and Wilbur Brookover. These researchers and others named the genre "Effective Schools Research"; over time this body of research has been through four phases. According to Lezotte and Snyder (2011) they are:

1) Finding effective schools (1966-1973)[3]

2) Descriptive phase (1973-1980)

3) Prescriptive phase (1980-1990)

4) Total systems alignment phase (1990-present)

In the beginning, researchers set out first to find effective schools, defined as schools that made a difference in student achievement, and were serving students from low-income families. This was followed by a descriptive phase where researchers described the inner workings of the school, a prescriptive phase where the effective schools correlates were shared, a school district phase where broad-based organizational support was identified, and a total system alignment phase where every effort in the school system was aligned to goals and effective schools

[3] All time stamps are based upon my perspective after reviewing the effective schools literature. They do not necessarily represent the perspective of Lezotte and Snyder.

correlates (Lezzotte & Snyder, 2011). Reynolds, Teddlie, Creemers, Scheerens, and Townsend (2000) describe the stages of school effectiveness research slightly differently. They also identified four stages:

> Stage 1, from the mid- 1960s and up until the early 1970s, involved the initial input-output paradigm, which focused upon the potential impact of school human and physical resources upon outcomes;
>
> Stage 2, from the early to the late 1970s, saw the beginning of what were commonly called the "effective schools" studies, which added a wide range of school processes for study and additionally looked at a much wider range of school outcomes than the input-output studies in stage 1;
>
> Stage 3, from the late 1970s through the mid-1980s, saw the focus of SER shift towards the attempted incorporation of the effective schools 'correlates' into schools through the generation of various school improvement programmes;
>
> Stage 4, from the late 1980s to the present day, has involved the introduction of context factors and of more sophisticated methodologies, which have had an enhancing effect upon the quality of all three strands of SER (school effects research, effective schools research, school improvement research) (p. 4).

Reynolds, Stringfield, Teddlie, and Creemers (2002) describe the effective schools research history as consisting of three generations or periods of the movement:

> Generation 1- Finding/Identifying positive and negative outlier schools (effective versus ineffective schools).
>
> Generation 2- The use of advanced statistical techniques are used for projects and research including the use of hierarchical linear modeling (HLM).
>
> Generation 3- An understanding of context variables is developed and as a result, school effectiveness research plateaued.

In addition, Reynolds et al. (2000) describe three major strands of school effectiveness research:

- School Effects Research- studies of the scientific properties of school effects evolving

from input-output studies to current research utilizing multilevel models;
- Effective Schools Research- research concerned with the process of effective schooling, evolving from case studies of outlier schools through to contemporary studies merging qualitative and quantitative methods in the simultaneous study of classrooms and schools;
- School Improvement Research- examining the processes whereby schools can be changed utilizing increasingly sophisticated models that have gone beyond simple applications of school effectiveness knowledge to sophisticated 'multiple lever' models (p. 3).

While these phases and strands are not universally used within the school effectiveness literature, they are useful to this literature review as they help to define the topic. It is clear from the progression in the research and from these phases and strands that researchers initially set out to find effective schools, then identify what made them effective, then determine how to implement those effective strategies in all schools to ensure student learning. Even though this research has morphed, schools today are still trying to implement best practices in effective schooling. This is evidenced by the way school improvement plans across the United States are developed and the items they include. The current phase is still in progress.

Finding Effective Schools

After the release of the Coleman Report, researchers began looking for evidence of schools that were effective for poor children. Researchers at the time called this the "Search for Effective Schools Project" (Edmonds & Fredericksen, 1979, p. 4). Lezotte, Edmonds, and Ratner (1974) analyzed all of the elementary schools in the Detroit, Michigan Model Cities Neighborhoods. Detroit was one of 139 cities that was selected and awarded funds to be a part of the Department of Housing and Urban Development's Model Cities Program. Cities were identified based upon their struggles with poverty. The program was developed to demonstrate that the welfare of people living in "slum and blighted neighborhoods could be improved substantially through concentration of federal, state, and local efforts" (Samuelson, 1972, p. 1).

Between 1966 and April of 1971, $704 million was awarded through the program.

The school effectiveness researchers chose the schools in Detroit to study because they could control for income and social class, due to the availability of demographic data through the Model Cities Program. In addition, student achievement data was available as well from the Iowa Test of Basic Skills and the Stanford Achievement Test. Lezotte, Edmonds, & Ratner (1974) found evidence within the twenty schools that were analyzed, that factors within the school did appear to influence student achievement after controlling for income and social class (Lezotte, Edmonds, & Ratner, 1974; Lezzote & Passalacqua, 1978; Edmonds & Fredericksen, 1979; Edmonds, 1980).

Klitgaard and Hall (1974) performed a rigorous large-scale study over a two-year period. They investigated 4th and 7th grade scores from ninety percent of the schools in Michigan, 2nd through 6th grade scores in New York City, and scores from the Project Talent high school study. They found that fifteen Michigan schools and thirty New York City schools increased student achievement by one standard deviation above the mean after controlling for background variables (Marzano, 2000).

Realizing that this small number of schools alone would not make a strong enough case that schools could overcome the barriers that poverty presented to critics, researchers continued to look for schools that were serving predominantly poor students and had come very close to eliminating the effects that poverty had on student achievement (Edmonds, 1980). After working in Michigan due to accessibility to necessary data and the Northeast corridor of the United States, the search broadened.

In 1978, researchers studied a random sample of elementary schools in Michigan in an attempt to validate earlier findings by Brookover and colleagues (1978). They found that while:

> school-level SES, racial composition, and climate were each highly related to mean school achievement; only a small portion of the between-school variance in achievement is explained by SES and racial composition after the effect of school climate is removed. The climate variable we have called Student Sense of Academic Futility had the largest correlation with achievement (Brookover, Schweitzer, Schneider, Beady, Flood, & Wisenbaker, 1978, p. 301).

The findings of this work were important because they showed that the school social environment and aspects of that environment clearly had an impact on student achievement. They were also important because school climate was something that schools could control, shape, and continually develop. In addition, while SES and race clearly explained significant portions of the variance in school performance, they only explained a small portion of the between-school variance after controlling for the social-psychological climate variables. In their conclusion, the authors stated,

> predominantly low SES and minority schools are more likely to be characterized by a high sense of student academic futility, low academic norms and low expectations for and evaluations of the students. In fact, these composition characteristics frequently may contribute to the development of differential expectations, norms, and feelings of futility. (Brookover et al., 1978, p. 316)

This was the early evidence that a "Culture of High Expectations" was important. This later became the second correlate of Effective Schools.

Another early large-scale study was The Louisiana School Effectiveness Study (Stringfield & Teddlie, 1988). The study was designed in response to the state legislature's directive to the state department of education to investigate the conditions related to student achievement in Louisiana elementary schools (Stringfield & Teddlie, 1988). The researchers intended to conduct the study in five phases: the first, which began in 1981, was a series of questionnaires based upon the work of Brookover. Phase two included the selection of a random stratified sample of seventy-six elementary schools that then participated in questionnaires

designed to assess school climate and other variables (Stringfield & Teddlie, 1988).

Phase three of the study included the identification of sixteen schools, eight outlier pairs, which were then researched on a more in depth basis. These pairs were from the same geographic region, their demographics were very similar, but their student achievement levels were very different. In this portion of the study, Stringfield and Teddlie (1988) found that 13% of the achievement variance between schools could be attributed to conditions within individual schools, while 12% could be attributed to differences between teachers within schools. The researchers added,

> Differences in schools have a slightly greater effect on student achievement than differences in teachers; both these effects show that schools and teachers, not just socioeconomic status of students, do strongly influence achievement (Stringfield & Teddlie, 1988, p. 44).

Classroom observations also occurred in the outlier pairs, in both the more effective and less effective schools in terms of student achievement during phase three. The instructional findings in effective schools are worth noting. They were characterized by more time on task by students, much more interactive instruction, and a higher usage of direct instruction being utilized (Stringfield & Teddlie, 1988).

As evidenced by the previously highlighted research, it did not take long for conclusions to be reached from the counter studies to the Coleman Report. Many researchers studying the effects of school on student achievement had come to the conclusion that schools do have a significant impact on student achievement (see for example, Chubb & Moe, 1990). In 1979, one of the most prominent researchers in the effective schools movement had formed an important deduction based on all of the research that followed the Coleman report. Ronald Edmonds said,

> We can, whenever and wherever we choose, successfully teach all children whose schooling is of interest to us. We already know more than we need to do that. Whether

> or not we do it must finally depend on how we feel about the fact that we haven't so far (Edmonds, 1979b, p.23).

This statement has echoed through many public schools throughout the nation as educators seek to validate their beliefs, efforts, and values. As these beliefs, efforts, and values became more solidified, the mission of public schools seemed to begin to change as well. As Lezotte (1997) pointed out, the new mission for schools is to do whatever it takes to ensure learning for all. In other words, it is the school's job to ensure high levels of learning for all students.

Effective Schools Literature

The effective schools literature base is very large. Searching "effective schools" in EBSCO host yields 2,378 results[4]. These studies were published between 1964 and 2013. Because the bulk of what was considered the initial "effective schools research" was done in the 1970's and 1980's, there was a need to evaluate additional research. This requires different search parameters. Other key terms that emerged from the research that contribute to what I define as an effective school include: school effects, school effectiveness, and comprehensive school reform. Table 2.1 shows how the research morphed throughout different time periods.[4] As evidenced by the table, category names became more solidified over time with "school effectiveness" being the most popular term to describe school level positive results in student achievement.

[4] Data retrieved from ERIC-EBSCO Host on March 23, 2013

Table 2.1: Search Terms Over Time

Search Terms	1964-1980	1981-1990	1991-2000	2001-Present
Effective Schools	**68**	**866**	**640**	**804**
School Effects	**79**	**72**	**101**	**136**
School Effectiveness	**107**	**2688**	**2382**	**1961**
Comprehensive School Reform	**1**	**1**	**112**	**345**

Teddlie and Stringfield (2007) describe school effectiveness research (SER) in the United States as consisting of "two closely linked branches" (p. 137). They are effective schools research and school effects research. They distinguish between the two branches in the following way:

- *Effective schools research*. This research is concerned with the processes of effective schooling and, in its initial phases, involved the generation of case studies of positive outlier schools that produced high achievement scores for students living in poverty. Cumulative results from effective schools research have resulted in detailed descriptions of effective school characteristics across a variety of contexts. The best-known findings from SER come from these studies.

- *School effects research*. This research involves the study of the scientific properties of school effects (e.g., the existence and magnitude of school effects, the consistency and stability of school effects). The initial studies involved the estimation of the impact of schooling on achievement through the regression-based input-output studies in economics and sociology. This branch of SER has always placed an emphasis on methodological issues, which has become a hallmark of the tradition (e.g., Teddlie, Reynolds, & Sammons, 2000)

Effective schools research focuses on educational processes, while school effects research focuses on educational products (2007, p. 137).

This distinction is helpful when reviewing effective schools literature. Especially when interpreting the differences between case studies done with outlier schools from effective schools research and product-based studies like those done by Coleman and colleagues.

As some researchers have pointed out, it seemed apparent that one of the purposes of effective schools research was to develop practical means for school improvement (Townsend, 2007). It is important to understand the similarities and differences with effective schools research and school improvement research. There has been a very close relationship between the two over the years. One key difference according to Smink (1991) is that the focus of school effectiveness is on results. Measurable variables are identified that lead to school success. The focus of school improvement is on the process. A broad list of variables that are part of the school improvement process are utilized and the plan is built based upon them.

Original Effective Schools Correlates

A seminal article called, *Effective Schools for the Urban Poor* (Edmonds, 1979b) first framed the effective schools correlates. This is an article that has been cited multiple times in the Effective Schools Research literature and is considered to be one of the most influential articles from effective schools research.[5] Edmonds's (1979b) article was important because it summarized much of the research that had been done by the early effective schools researchers. He also discussed the search for effective schools and the findings that were emerging. What was probably the most ground breaking about this article was that it provided the "Effective Schools Characteristics" and highlighted each of them.

[5] In a search of the Web of Science Database, this article was cited in other articles 364 times. This is more than triple the next most cited article on the topic.

Five of the six Edmonds listed in his summary later became known as the Effective Schools Correlates, or those factors that appear in effective schools. Those listed in 1979 were:

1) Strong Administrative Leadership

2) Climate of High Expectations

3) Orderly School Atmosphere

4) Basic School Skills Take Precedence Over Other Activities

5) School Energy and Resources can be Diverted from other Business to Fundamental Objectives

6) Progress is Frequently Monitored

The sixth characteristic listed by Edmonds reinforces the other five, but did not remain on the conventional list of effective schools correlates as time progressed through the 1980's. As one of the most prominent contributors to the effective schools movement until his death in 1983, Edmonds framed many of the concepts of the movement along with the correlates (Edmonds, 1979a, 1979b, 1979c, 1980, 1981).

Additional Early Effective Schools Correlates

In addition to the original list of effective schools correlates provided by Edmonds, other lists existed as well (Purkey & Smith, 1983). Table 2.2, taken from Scheerens and Bosker (1997, p. 154) illustrates many of the early case studies of Effective Schools and the factors identified in the case studies as influencing achievement, as highlighted by Purkey and Smith (1983).

Table 2.2: As reported by Sheerens and Bosker (1997, pg 154)

Author	Weber (1971)	Venezky & Winfield (1979)	Glenn (1981)	Cal. State (1980)	Brookover & Lezotte (1979)	Brookover *et. al.* (1979)	Rutter *et. al.* (1979)
No. of Schools	4	2	4	--	8	159	12
Strong Leadership	x		x	x		x	
Orderly Climate	x		x				x
High Expectations	x		x	x	x	x	x
Frequent Evaluation	x		x	x			
Achievement-Oriented Policy		x		x		x	x
Cooperative Atmosphere		x	x				x
Clear goals for Basic Skills			x		x		
In-service Training/ Staff Development		x		x			
Time on task					x	x	x
Reinforcement						x	x
Streaming						x	x

After investigating the case studies in table 2.2, Purkey and Smith concluded that five factors showed up in most of the case studies. They were:

> strong leadership by the principal or other staff, high expectations by staff for student achievement, a clear set of goals and an emphasis for the school, an effective schoolwide staff training program, and a system for the monitoring of student progress (1983, p. 435).

These factors were prevalent in the early literature, including the list of correlates put forth by Edmonds (1979). The school-wide staff-training program was a noticeable addition to the list Edmonds put forth. Sheerens and Bosker (1997) pointed out that there was much consistency

between the lists. The researchers also indicated that there seemed to be consensus on many of the items noting that strong leadership was mentioned four times, orderly climate three times, high expectation six times, achievement oriented policy four times, and time on task three times.

While Purkey and Smith (1983) compiled the Effective Schools variables and noted there was some consistency in the results from many studies, they also warned, "the variations in findings should serve as caution to those who would reduce such disparate literature to five or six variables" (p. 431). Additional critiques will be discussed later in this review.

While many of the early studies on school effects found different variables or correlates to be important, over time, the larger body of research on effective schools morphed. Specifically, it included seven correlates rather than the original five characteristics outlined by Edmonds in 1979, and others that were included in additional studies. These were: high expectations for success, instructional leadership, clear and focused mission, opportunity to learn/time on task, frequent monitoring of student progress, safe and orderly environment, positive home-school relations (Lezzotte & Snyder, 2011; Levine & Lezotte, 1990; Lezotte & Pepperl, 1999).

Effective Schools Process for Continuous School Improvement

Once the correlates were identified, there was a shift in many schools to try to implement the correlates as a school improvement process. While environmental factors beyond the influence of the school undoubtedly made the mission of learning for all more challenging, the prevailing view from effective schools researchers was that the school had control over enough variables to ensure that virtually all students learn (Lezotte, 1997). With the correlates identified, and this belief serving as a driving force, momentum for change in schools began with an effort to put the correlates into place in schools. This led to the development of The National

Center for Effective Schools (NCES) in 1986 (Rossmiller & Halcomb, 1993). As Rossmiller and Halcomb put it, "The correlates can be characterized as a list of ingredients that are required for school effectiveness, but they are not a recipe" (1993, p. 4).

In the late 1980's, NCES initiated a school improvement and staff development program based upon the Effective Schools Model (1993). Schools were able to use this as a road map to becoming an effective school. Many schools and school districts were able to implement the school improvement model developed by the National Center for Effective Schools with successful results. Some of these schools and school districts have been researched.

One example was the Clarke County Public School District in Clarke County, Virginia. Clarke County Schools began their journey in 1989 and, after four years, received the U.S. Senate's Productivity Award, which is issued to only one public agency each year (Jones & Ross, 1994). Jones and Ross analyzed achievement results for all students in the district as a case study for Effective Schools as a school improvement model. Over a four year period from 1990 to 1993 Clark County Reading passing percentages on the Virginia Literacy Passport Test in grade 6 increased from 55 percent to 79 percent while the statewide average went from 65 percent to 69 percent during the same timeframe. Results for writing were less dramatic, but the district still outgained the state (Jones & Ross, 1994).

In their conclusion, Jones and Ross wrote, "No recipe can be provided for the redesign of public education, no quick fix is available- but there is a framework for school improvement in the Effective Schools Research model." (p. 22). After looking at much of the effective schools literature, it is remarkable how consistent the effective schools framework has been over the course of time. While many have provided a list of variables that influence student achievement, the lists are all pretty consistent.

Effective Schools Research Methods

The original research on effective schools considered many variables that related to student achievement and the correlates emerged. This was a result of first identifying effective schools and then researching them to determine what practices they implemented that led to increased student achievement. Due to this factor, the correlates were rarely studied in isolation early in the effective schools movement. This was largely because there were multiple practices that appeared to influence student achievement.

Schools that were studied due to the effectiveness they demonstrated in terms of student achievement were doing many things that contributed to those higher levels of student achievement. Isolating practices and studying their impact on student achievement was not possible due to the nature of the methods (first finding the effective schools and then studying what they were doing to generate high levels of achievement). This creates a situation where the variables are confounded. In addition, schools are complex organizations. Therefore, what works in one school and the unique context that is associated with it does not necessarily work in another school with a different context. This presents many challenges in effective schools research.

Critiques of Effective Schools Research

Over the years, there have been many researchers who have raised concerns regarding Effective Schools Research. They point out the fact that the research is somewhat weak due to the often-narrow recipes for school improvement inherent in the Effective Schools Correlates (Coe & Fitz-Gibbon, 1998; Purkey & Smith, 1983). Others suggest that what is really happening in schools deemed to be effective is that they are teaching to the test, narrowing their curriculum, and spending time on test taking skills (Stedman, 1985). Coe and Fitz-Gibbon (1998) point out

seven criticisms of school effectiveness research (SER). They are:

1. Validity of school effectiveness measures depends upon the outcomes that are being measured. Due to the data available, school effectiveness research has typically taken a fairly constricted range of outcomes and assumed that these outcomes (sometimes only one) are fitting for all students and schools.
2. The models used to estimate the effect the school has on student achievement have not addressed the construct of effectiveness.
3. A causal relationship between schools and the school's effects on student learning has not been demonstrated. Factors inside the school could be causing the differences in school effectiveness.
4. To date, there is no research that indicates that schools can have an impact on their effectiveness. It is unclear the extent to which a school can influence student outcomes by actions within the control of the school.
5. School effects have not been separated from classroom effects or individual teacher effects. This is poor statistical modeling.
6. The identification and then emphasis on the correlates of effective schools has slowed progress toward understanding the true nature of school effectiveness due to imprecise thinking about correlates.
7. The congruence between goals and measurement (goal-measurement congruence) could illuminate more than the designation of effectiveness in the research.

These criticisms certainly have some validity and need to be addressed. As I will highlight in the rest of this review, many of them have been fully addressed by the field in the years since 1998.

The research community has not been at consensus on the best methods to use in

identifying effective schools. Some researchers made the argument that in order to progress methodologically in the studies of school and classroom effects, methods of analysis for multilevel data would have to be developed (Burstein, 1980; Bryk & Raudenbush, 1988; Scheerens & Bosker 1997). In school effectiveness research, multilevel modeling (or hierarchical linear modeling) is superior to OLS regression models because of the various macro and micro variables in multilevel data like that of school effectiveness research (i.e. teacher effects, classroom effects, school effects, district effects, etc.). For a detailed explanation of multilevel modeling, see Snijders and Bosker (2012). In 1988, due to progress made in methods to analyze multilevel data, Bryk and Raudenbush concluded that,

> It is now possible for investigators to formulate and test explicit models describing how policies, practices, resources, and norms evident at each level of educational organization influence the distribution of outcomes at the next lower level (Bryk & Raudenbush, 1988, p. 468).

Sheerens and Bosker (1997) seem to corroborate this view, noting that there was a "lack of appropriate statistical procedures available to deal with the hierarchical nature of the data" (p. 72). Additionally, they cite the work of Aitkin and Longford pointing out that prior to 1986, studies "had to rely on (in hindsight) unsound statistical procedures" (p. 72).

Due to these insights, it is important to consider studies produced after 1986. It is also worth pointing out that this renders the conclusions from the Coleman Report and later the Jenks and colleagues study insufficient for the same reasons. The capability to interpret multilevel data makes the more recent effective schools studies a closer approximation of what is really happening in schools from a statistical point of view. Being able to distinguish between classroom effects, school level effects, and district level effects is important if the goal is to determine the effects a school has on student achievement as is the case in school effectiveness

research.

Is the School Effect Large Enough to Matter?

The effective schools research generated a substantial body of evidence that schools could make a difference for students, especially those from less privileged backgrounds. Another important question to answer is how much do schools matter? In other words, how much of the variance in student achievement can be explained by the school. Robert Marzano provided a response to this question in his paper titled, *A New Era of School Reform: Going Where the Research Takes Us* (2000).

Marzano synthesized the data from ten[6] different Effective Schools Studies to attempt to find the magnitude of impact the school had on student achievement. His findings were noteworthy because the percentage of variance explained by differences among schools was large enough to demonstrate that schools had a profound effect on student learning. The average standardized mean difference effect size (ESd) was .96, the average percentage variance (PV) explained by the school effect was 18.49% and the average percentile (P) gain associated was 33% (2000). There were outliers among the ten studies. Using the Q statistic as a measure of homogeneity, distributing it as chi-square with (k-1) degrees of freedom where k is the number of effect sizes in the set, and identifying elements that were statistically significant identified the outliers.

Put another way, as eloquently stated by Marzano:

---

[6] Studies included were: Coleman et al. (1966), Jencks et al. (1972), Bryk & Raudenbush (1992), Sheerens & Bosker (1997), Rowe & Hill (1994), Creemers (1994), Stringfield & Teddlie (1989), Bosker (1992), Luyten (1994), and Madaus et al. (1979).

> When the PV of schools is assumed to be 20.00, it implies that the percentage of students who would pass a state-level test (for which the expected passing rate is 50 percent) is 72.36 percent for effective schools versus 27.67 percent for ineffective schools, for a differential of 44.72 percent. This is not a trivial difference, especially for the 44.72 percent of students (2000, pg. 47)[7].

Table 2.3 provides a visual display of the data from the Marzano (2000) study.

Table 2.3: Magnitude of the School Effect adapted from Marzano (2000, p. 46)

Studies Included	ESd	Percentage Gain	Percentage Variance
All 10 Studies	.96	33%	18.49%
Outliers Removed	1.01	34%	20%

Marzano's (2000) focused on measuring whether or not the school had an impact on student achievement, and if it did, the magnitude of the impact that the school had on student achievement. This particular work demonstrated that the school did in fact have a fairly substantial impact on student learning. As highlighted above, the school is responsible for about 20 percent of the variance in student achievement.

In addition to Marzano's work, Bosker and Witziers performed a meta-analysis using 89 international studies on the gross school effects on student achievement (Scheerens & Bosker, 1997). Their findings indicated less of an impact by the school with a gross school effect of .48 and net effects at a magnitude of .30. Reporting on these results, Sheerens and Bosker note that, "schools account for 19% of the achievement differences between students, and for 8% when

[7] The reader might note that this is a binomial effect size display. In this type of statistical model, half the students would pass and half would fail. This is simply a model and does not mirror what actually happened in schools.

adjusting for initial differences between students" (1997, p. 77). Using Cohen's effect sizes, these are considered to be moderate effect sizes.

It should be noted that Bosker and Witziers meta-analysis on gross school effects included studies from multiple countries and multiple schools within each country. Scheerens and Bosker (1997) point out that studies from North America produced significantly higher effect sizes than other countries included in the meta-analysis. It is probable that some of the difference in the magnitude reported in this meta-analysis and the synthesis done by Bosker and Witziers is due to the inclusion of studies from many other countries with lower effect sizes reported whereas Marzano only looked at studies of U.S. schools. Also, schools have different impacts due to cultural contexts. This is particularly true when looking at the effects in schools in different countries because the cultural contexts vary greatly from one country to another.

This evidence clearly demonstrates that the schools effects on student achievement are large enough to warrant the time, effort, and energy of school personnel to work to ensure school effectiveness. In addition, this is further evidence that a meta-analysis focused on results in the United States in the area of effective schools is merited.

The First Effective Schools Meta-Analysis

In addition to their thorough review of the literature and the history of the effective schools movement, Sheerens and Bosker (1997) performed an effective schools meta-analysis and published the results in their book titled, The Foundations of Educational Effectiveness. This was not a comprehensive meta-analysis; rather, it was a sample meta-analysis, meaning the researchers only selected a sample of all of the studies on the topic to include in the meta-analysis. While it is not clear how many studies were included to compute the effect size of each of the variables addressed in the meta-analysis, the authors did indicate that for the

"Cooperation" variable, twenty studies were included. This meta-analysis was international in nature, but it did include studies from the United States. This meta-analysis included the following ten variables:

1) Cooperation (n= 20)[8]
2) School Climate (n= 22)
3) Monitoring measured at: (a) school level (n= 24); (b) class level (n= 43)
4) Opportunity to Learn, which concept is indicated by: (a) content coverage (n= 19); (b) homework (n= 13); (c) time (n= 21)
5) Parental involvement (n= 14)
6) Pressure to achieve (n= 26)
7) School leadership (n= 38)

The effect sizes for each variable are listed in Table 2.4. Generally speaking, these effect sizes are small. It is important to note the differences between the international effect size and the U.S.A. effect sizes. As Sheerens and Bosker point out in their work, the effects for achievement pressure and cooperation are higher for the U.S.A. (Sheerens & Bosker, 1997). Additionally, school climate and monitoring at the classroom level indicate negative effects in the U.S.A. The school climate variable included twenty-two studies, and the monitoring at the classroom level variable included forty-three studies. This could be considered a limitation to drawing conclusions about U.S. schools since Sheerens and Bosker did not report how many of those studies pertained to schools in the U.S.A. This is true for all of the variables included in this meta-analysis. This is more evidence that a more current meta-analysis should be performed based upon data from the United States only.

[8] n= the number of studies included in the meta-analysis for each variable

Table 2.4: School Level Effectiveness-Enhancing Factors; Meta-analysis (adapted from Sheerens and Bosker, 1997)

Variable	Mean Effect Size	S.E.	U.S.A. Effect	S.E.
Cooperation	0.0292	0.0185	0.0798	0.0579
School Climate	0.1090	0.0333	-0.1442	0.0918
School Level Monitoring	0.1481	0.0370	0.0462	0.3218
Class Level Monitoring	0.1147	0.0822	-0.3745	0.4543
Content Coverage	0.0886	0.0386	0.1638	0.1126
Homework	0.0574	0.0260	0.1100	0.1218
Time	0.1931	0.0491	0.2163	0.1306
Parental Involvement	0.1269	0.0596	0.2634	0.2696
Pressure to Achieve	0.1327	0.0441	0.2499	0.1097
School Leadership	0.0499	0.0225	0.0169	0.0699

All Effect Sizes Reported have been transformed to Fisher's Z (Z_r)

Recent Effective Schools Progress, Correlates Updated

While Edmonds' original five correlates of effective schooling served as the anchor for the early part of the effective schools movement, as pointed out in this review, as more and more research accumulated in this area, expansion from the original five and then seven correlates did occur. Levine and Lezotte (1990) summarized the effective schools research and generated nine characteristics. Marzano (2003) provided a synthesis of effective schools variables that included five school level factors that impact student achievement. They were:

1) Guaranteed and Viable Curriculum

2) Challenging Goals and Effective Feedback
3) Parent and Community Involvement
4) Safe and Orderly Environment
5) Collegiality and Professionalism

The definition each researcher associates with each item on their list is important because it provides additional clarity. For example, Marzano (2003) indicates that the variable, Challenging Goals and Effective Feedback includes setting high student goals and providing them feedback in relationship to those goals, but it also includes high expectations for student learning that appears on so many effective schools correlate lists. In addition, the factor of Collegiality and Professionalism included staff development.

Other researchers did the same type of synthesis of correlates as well, including Reynolds and Teddlie (2000) and Teddlie and Stringfield (2007). The most recent list found (Teddlie and Stringfield, 2007)[9] includes nine effective schools process variables and several subcomponents of each process variable. The nine variables and subcomponents were:

1) The processes of effective leadership
    a. Being firm and purposeful
    b. Involving others in the process
    c. Exhibiting instructional leadership
    d. Frequent, personal monitoring
    e. Selecting and replacing staff
2) Developing and maintaining a pervasive focus on learning
    a. Focusing on academics

[9] As noted by Teddlie and Stringfield (2007, p. 143) the process variables were adapted from Reynolds and Teddlie (2000, p. 144) which was a list developed by extracting the common elements from two other reviews (a) Levine and Lezotte (1990), and (b) Sammons, Hillman, and Mortimore (1995).

    b. Maximizing school learning time
3) Producing a positive culture
    a. Creating a shared vision
    b. Creating an orderly environment
    c. Emphasizing positive reinforcement
4) Creating high (and appropriate) expectations for all
    a. For students
    b. For staff
5) Monitoring progress at all levels
    a. At the school level
    b. At the classroom level
    c. At the student level
6) The process of effective teaching
    a. Maximizing class time
    b. Successful grouping and organization
    c. Exhibiting best teaching practices
    d. Adapting practice to the particulars of classroom
7) Involving parents in productive and appropriate ways
    a. Buffering negative influences
    b. Encouraging productive interactions with parents
8) Developing staff skills at the school site
    a. Site based
    b. Integrated with ongoing professional development
9) Emphasizing student responsibilities and rights
    a. Responsibilities
    b. Rights

These served as the basis of my coding scheme for the research synthesis/meta-analysis. I chose to use these nine variables because they represent the most recent updated combined list of correlates and they were presented by some of the most prominent effective schools researchers of our time, Reynolds and Teddlie, and adapted by Teddlie and Stringfield. Additionally, the list was not presented as a recipe, rather, it was reproduced as a table by Teddlie and Stringfield (2007, p. 143) to serve four key purposes:

(1) Serve as a visual representation of how the original correlates expanded, now including nine processes.

(2) It shows how the nine processes are much more intricate than the correlates

originally reported by Edmonds. Subcomponents of each process variable are included. This encompasses the vast expansion of school effectiveness research and includes findings from various school contexts.

(3) It reveals that research from other areas is captured, including teacher effectiveness research.

(4) It shows the sheer scale of school effectiveness research. As an example, it includes the Teddlie and Reynolds (2000) study with over 1400 references.

This list was used because it was a combination of all of the effective schools research variables. It morphed from the original five correlates (Edmonds, 1979b) to the nine listed. It also addresses some of the criticisms of effective schools research, namely that it has been too simplistic and lacked multi-level analysis. Because this list includes variables deemed prominent from teacher effectiveness research (TER), and subcomponents include individual student level subcomponents, the list is more representative of the current understandings of effective school correlates. For the purposes of this research synthesis, the focus will be on creating high and appropriate expectations for all and developing staff skills at the school site.

While both of these areas have been studied in some detail, neither have been studied with the lens of effective schools in mind. Probably the most comprehensive work on these two topics was done by John Hattie (2009, 2012) in his meta-analysis of now over 900 meta-analyses. Hattie found that both professional development and teacher expectations had a positive effect. The effect sizes (or magnitude of the effect) were .51 for professional development and .43 for teacher expectations (Hattie, 2012). Both of these effect sizes are considered to be in the moderate to high range.

Hattie's (2012) meta-analysis of meta-analyses included 10 meta-analyses on

professional development[10]. The effect size of .51 was not the effect professional development had on student achievement, rather, it was a combined effect size of changes such as: the changes in teacher learning, the changes in teacher behavior, teacher reactions to the professional development, teacher skills, and changes in student achievement. As Hattie (2009) stated, "One of the difficulties with reviews of professional development is that the outcomes seem to be more about changes in the teachers, and not about the impact of professional development on student outcomes" (p. 119).

This is a very important distinction between Hattie's work and the work that has been done in this meta-analysis, because the sole focus in relationship to professional development in this meta-analysis is on student achievement results. It is also important to mention that Hattie's analysis included studies from all over the world, and was not focused on school-wide results, but all results related to professional development.

Hattie's (2012) meta-analysis of meta-analyses in the area of teacher expectations included 8 meta-analyses[11]. While the effect size for teacher expectation was .43, this also included meta-analyses with studies from all over the world. Additionally, it was not focused on school-wide results, which is a contrast to the work done in this meta-analysis.

---

[10] The ten meta-analyses included: Joslin (1980), Harrison (1980), Wade (1985), Blank & Alas (2010), Lomos, Hofman & Bosker (2011), Salinas (2010), Yoon, Duncan, Lee & Shapley (2008), Batts (1988), Tinoca (2004), and Temperley, Wilson, Barrar & Fung (2007).

[11] The eight meta-analyses included: Rosenthan & Rubin (1978), Smith (1980), Dusek & Joseph (1983), Raudenbush (1984), Harris & Rosenthal (1985), Ritts, Patterson & Tubbs (1992), Ide, Parkerson, Haertel & Walberg (1995), and Tenebaum & Ruck (2007).

Summary

Effective schools research has been ongoing for almost forty-five years, since one of the initial Silberman (1970) studies appeared. The initial research was a response to the Colman report and Jencks et al. report that schools did not make a difference for students in terms of their achievement levels. The effective schools research has grown and become much more refined as time has passed.

We now know that schools do in fact have an effect on student learning and we have at least two studies that quantify that effect (Scheerens & Bosker, 1997; Marzano 2000). We also know that there is widespread agreement in the research community that there are some common indicators that influence student achievement. The focus of my research is to build on this by trying to quantify the magnitude of the effect high and appropriate expectations for students and developing staff skills at the school site have on student achievement at the school level by performing a meta-analysis. While effect sizes and meta-analyses already exists for both of these topics, to date there has not been a meta-analysis in these areas that only considered these topics through the lens of effective schools, or school-wide implementation and results in terms of student achievement. These two areas were chosen over other effective schools correlates based upon the speculation that they might generate higher effects on student achievement than others.

Chapter 3

Method

To respond to my research questions, I used meta-analysis. Meta-analysis is a quantitative research synthesis technique (Lipsey & Wilson, 2001). The purpose of a research synthesis/meta-analysis is to compile empirical studies done on a topic, analyze the information, and draw conclusions from the compiled results of previous studies (Cooper, 2010). According to Lipsey and Wilson (2001) there are four main advantages of meta-analysis.

1) Meta-analysis provides a scientific way to summarize the results from multiple research findings.
2) Meta-analysis provides the tools necessary to distinguish between key study findings in a more sophisticated way than typical qualitative summaries.
3) Meta-analysis allows the researcher to find relationships that are not highlighted in other summarizing approaches.
4) Meta-analysis offers a way to compile and organize information from a multitude of studies within a single review.

As highlighted in table 2.1 in chapter 2 of this dissertation, there are many studies that have been done on effective schools. There are at least 9,516 studies that use the terms "effective schools" or "school effects" alone[12]. This renders this particular topic excellent for a meta-analysis. Since so many studies have been done, it was likely that data from studies can be combined and a meta-analysis performed.

[12] Data pulled from ERIC-EBSCO Host on March 23, 2013

Research Questions Restated

There are many variables that have an impact on the effectiveness or ineffectiveness of a school in relationship to student academic performance. For this study, the first area of focus was on measuring the impact that expectations of students and expectations of staff have on student achievement. In other words, to what degree do expectations impact student achievement?

The second area of focus was on the impact that staff development has on student achievement. For an additional area of focus, please see Prenkert (2015). Together these two studies represent four of the nine effective schools variables highlighted in chapter two.

Procedures

There are seven steps to conducting a meta-analysis as well as several questions concerning the validity of conclusions. They are listed below in Table 3.1 as presented by Cooper (2010, pp 18-19).

Table 3.1: A Checklist of Questions Concerning the Validity of Research Synthesis Conclusions

Step 1: Formulating the problem
1. Are the variables of interest given clear conceptual definitions?
2. Do the operations that empirically define each variable of interest correspond to the variable's conceptual definition?
3. Is the problem stated so that the research designs and evidence needed to address it can be specified clearly?
4. Is the problem placed in a meaningful theoretical, historical, and/or practical context?
Step 2: Searching the literature
5. Were proper and exhaustive terms used in searches and queries of reference databases and research registries?
6. Were complimentary searching strategies used to find relevant studies?
Step 3: Gathering information from studies
7. Were procedures employed to ensure unbiased and reliable (a) application of criteria to determine the substantive relevance of studies and (b) retrieval of information from study reports?
Step 4: Evaluating the quality of studies

8. If studies were excluded from the synthesis because of design and implementation considerations, were these considerations (a) explicitly and operationally defined and (b) consistently applied to all studies?
9. Were studies categorized so that important distinctions could be made among them regarding their research design and implementation?

Step 5: Analyzing and integrating the outcomes of studies

10. Was an appropriate method used to combine and compare results across studies?
11. If a meta-analysis was performed, was an appropriate effect size metric used?
12. If a meta-analysis was performed, (a) were average effect sizes and confidence intervals reported and (b) was an appropriate model used to estimate the independent effects and the error in effect sizes?
13. If a meta-analysis was performed, was the homogeneity of effect sizes tested?
14. Were (a) study design and implementation features (as suggested by Question 8 above) along with (b) other critical features of studies, including historical, theoretical, and practical variables (as suggested by Question 4 above) tested as potential moderators of study outcomes?

Step 6: Interpreting the evidence

15. Were analyses carried out that tested whether results were sensitive to statistical assumptions and, if so, were these analyses used to help interpret the evidence?
16. Did the research synthesists (a) discuss the extent of missing data in the evidence base and (b) examine its potential impact on the synthesis' findings?
17. Did the research synthesists discuss the generality and limitations of the synthesis' findings?
18. Did the synthesists make the appropriate distinction between study-generated and synthesis-generated evidence when interpreting the synthesis' results?
19. If a meta-analysis was performed, did the synthesists (a) contrast the magnitude of effects with other related effect sizes and/or (b) present practical interpretation of the significance of the effects?

Step 7: Presenting the results

20. Were the procedures and results of the research synthesis clearly and completely documented (pp 18-19)

The steps and questions in Table 3.1 served as a roadmap for my research synthesis. Step 1 and the questions listed in step 1 were addressed in the first two chapters of this dissertation. Steps 2 through 7 are addressed in the remainder of this work.

Searching the Literature

While the list of exhaustive search terms grew as the review of the literature for studies that met the eligibility criteria for this meta-analysis continued, Appendix A includes the final list of search terms. It is important to note that I received advice on the initial search terms from the Regional Educational Laboratory Midwest (REL Midwest). I contacted REL Midwest through the "ask a REL" feature that allows individuals seeking research resources on education topics to submit their request, question is then delivered to the researcher with the most expertise in the given area and a response is generated (REL Midwest, 2014).

Specifically, I asked REL Midwest what search terms would be recommended for finding articles that addressed the impact that staff professional development had on student achievement and the impact that holding high expectations for all students had on student achievement. The recommended search terms from REL Midwest are included in Table 3.2. The number of studies the search terms generated in the ERIC database from 2001 to present is also listed in Table 3.2. Column two contains the number of studies generated with the search terms listed. Column three includes the number of studies with the terms student achieve* or academic perform* replaced with the term effect size. For the purposes of this synthesis, the term effect size was used in all searches as a way to narrow the scope of studies to those that were most likely pertinent to the present meta-analysis. Both lists are presented, so the reader can see the differences in the number of studies found using different terms.

Table 3.2: Search Terms and Number of Studies Each Term Generates in ERIC[13]

Search Terms for Expectations	# of Studies	Term "Effect Size" Added
Teach* expect* AND student achieve*	385	8
Teach* expect* AND academic perform*	61	NA
Teach* standards AND student achieve*	372	17
Teach* standards AND academic perform*	35	NA
Instruction* expect* AND student achieve*	51	1
Instruction* standards AND academic perform*	20	8
Instruction* standards AND academic achieve*	138	NA
Instruction* goals AND student achieve*	92	8
Search Terms for Staff Development		
Professional development AND student achieve*	1638	85
Teach* train* AND student achieve*	421	25
Teach* professional development AND student achieve*	635	36
"In-service" AND Teach* or train* AND student achieve*	718	82
Principal PD OR Admin* PD AND student achieve*	287	246
Principal train* OR Admin* train* AND student achieve*	295	274
Teach* Professional development AND academic perform*	46	NA
Teach* train* AND academic perform*	49	NA
Principal PD OR Admin* PD AND academic perform*	246	NA
Principal train* OR Admin* train* AND academic perform*	272	NA

Table 3.2 highlights the scope of the literature review. There were 42 studies reviewed for teacher expectations. There were 748 studies reviewed for staff development. While there were duplicate studies included within the various searches, I still needed to thoroughly investigate all of them to ensure all studies were included in the meta-analysis that met the eligibility criteria.

[13] Data gathered from ERIC on March 30, 2014.

My main search strategy was the use of the Education Resources Information Center (ERIC) database available through Ball State University. In addition, I included studies that met my eligibility criteria that were referred to me by colleagues or dissertation committee members. While primarily using one database excluded my meta-analysis from being considered comprehensive in nature, it was necessary for my work to narrow the scope of this dissertation. My dissertation chair, Dr. Joseph McKinney agreed to this limitation in the study. One of the goals of many research syntheses is to do a comprehensive search of the literature (Cooper, 2010, Lipsey & Wilson, 2001). This requires searching multiple databases, the Internet, and any other sources that could contain valuable data for the synthesis. While I am committed to continue compiling data throughout time on this topic, for purposes of this dissertation I only searched the ERIC database. The ERIC database is a very good database, but this was one of the limitations of this study, and one that is fully acknowledged.

Where this Study Fits Into the Methodology of Meta-analysis

The terms research review, research synthesis, and systematic review are sometimes used interchangeably; however, for the purpose of this dissertation I used the term limited research synthesis to describe my effective schools research. According to Cooper (2010), the terms research review and systematic review can also be applied to the critical analysis of a single study. As stated throughout the methodology section of this dissertation, this study was comprised of more than one study. Effect sizes were pulled from various studies for the purpose of conducting a meta-analysis.

Not all meta-analyses are performed based upon the same standards of reviewing the literature and gathering studies. For example, the Institute of Education Sciences, a wing of the federal government now routinely awards grants for meta-analyses. For many of these, it is

expected that a comprehensive literature review will be done searching multiple databases and other sources. The present meta-analysis was much more limited than that since I only used the ERIC database, and I was only searching for studies that report an effect size. Clear guidelines for classifying meta-analysis have not yet been developed; therefore, there was a need to describe where my work fits into the methodology. I have chosen to use the term limited research review to describe where my dissertation fits into the methodology of meta-analysis.

Gathering Information from Studies

The most time intensive part of the meta-analysis was searching the literature and gathering information from the studies found. A coding scheme was developed in collaboration with my research team and with approval of the dissertation committee. The full coding scheme can be found in Appendix B, but it is important to note that the scheme was developed based upon the suggestions from Cooper (2010). The full coding scheme presented in Appendix B represents a base scheme that could be used for a comprehensive meta-analysis of all of the effective schools variables. For the purpose of this meta-analysis, in column 2 of the coding scheme, only information on expectations (code 4) and staff development (code 8) were gathered.

To expedite future effective schools research and possibly additional meta-analyses, I decided the coding scheme would be more fully developed and shared as if all effective schools variables were part of the meta-analysis. During information gathering, it was critical that a good eligibility criterion was in place and that it was adjusted as studies are found that are ambiguous when the criteria was applied (Lipsey & Wilson, 2001). As Lipsey and Wilson (2001) point out, “We should caution, however, that only rarely does the version of the eligibility criteria that is first drafted serve well without revision” (p. 19). The eligibility criteria for this

study was constantly refined throughout the project and was constantly considered while information was gathered from studies.

Eligibility Criteria

One of the original effective schools researchers, Lawrence Lezotte, now defines the effective school in this way:

> The effective school is characterized by high overall student achievement with no significant gaps in that achievement across the major subgroups in the student population (Lezotte & Snyder, 2011).

While I agree with this description of an effective school, there is the additional need to characterize an effective school study for purposes of this research synthesis. I defined an effective schools study as any research indicating school-wide effects on student achievement. All empirical studies that reported on school-wide results in relationship to the variables in my coding scheme were included in this study. In addition, for inclusion in this study, it was required that the research be written in English and originate from the United States from 2001 to the present. This means that only studies that have been done since the passage of No Child Left Behind are included. This criterion was chosen because NCLB had such a profound influence on American education. Additionally, to be useful to the meta-analysis, research had to report data that could be used to generate an effect size. By necessity, the studies I used are quantitative in nature and report effect size data.

Table 3.3 below represents my full eligibility criteria including distinguishing features, key variables, cultural and linguistic range, time frame, and publication types. This criteria template is based upon the example shared by Lipsey and Wilson (2001, pp. 20-21).

Table 3.3 Eligibility Criteria for Inclusion in Effective School Meta-Analysis

Distinguishing Features- Eligible studies must report on school-wide student achievement results. This is defined by reporting on an entire course or grade level of results and not those of individual classrooms. The student achievement results must be tied to one of the effective schools variables included in my coding scheme.
Key Variables- Studies must report at least one quantitative outcome measure of effective schools variables. Only studies from which an effect size can be computed are eligible.
Cultural and Linguistic Range- Studies must be conducted and reported in English and be analyzing a school from the United States.
Time Frame- Only studies conducted from 2001 to the present are eligible. (This restricts the studies to those since the passage of the No Child Left Behind Act and other significant changes in education).
Publication Types- Published and unpublished studies are eligible, including refereed journals, non-refereed journals, dissertations, government reports, technical reports, etc.

Evaluation of the Quality of Studies

Any study that was excluded from this meta-analysis was due to it not meeting eligibility criteria. No study was excluded based upon any other considerations. Further, all eligible studies were categorized and documented in the coding scheme. Appendix B represents the entire coding scheme for this meta-analysis. It evolved much like the eligibility criteria has evolved over time as more and more literature was gathered and adjustments were made where needed. The coding scheme was designed so that important distinctions could be made among the various studies included, and so that moderator variables can be considered in the analysis.

Double Coding Studies

Coding information from studies was a very important task in meta-analysis, and not one to be taken lightly. As Cooper (2010) put it, "coding for a research synthesis is not a one person job" (p. 101). Having a single person doing the coding is a huge job, which is one of the reasons

this meta-analysis only focused on two of the nine identified effective schools variables, but in addition, having a single person doing all of the coding increases the possibility of bias exponentially. For this reason, my colleague Debra Prenkert has also coded the most complex information. We then met and resolved discrepancies in the coding. The biggest coding issue addressed was how to combine multiple effect sizes that were reported out over multiple years. Studies that were double coded as well as discrepancies in the coding are shared in Appendix C.

Analyzing and Integrating the Outcomes of Studies

Comprehensive Meta-Analysis 2 Software (CMA2) was used to perform all statistical procedures for this study. Since the studies reported randomized mean difference effect sizes (Cohen's d and Hedges g), the effect sizes reported in the results section of this dissertation are Hedges g. The effect sizes are reported in chapter 4 along with confidence intervals, an estimate of the independent effects, and the error in effect sizes. Additionally, the homogeneity of effect sizes was tested and along with potential moderators of study outcomes. All of these results are shared in chapter 4. Essentially, a meta-regression was run across all coded studies.

Interpreting the Evidence

While results are shared and interpretation is done in chapters 4 and 5 of this dissertation, it is important to note a few methodological expectations for meta-analysis within this methods section. First, as stated earlier, Comprehensive Meta-Analysis Software was utilized for this research synthesis. All results were tested for sensitivity to statistical assumptions. Missing data, is discussed and examined along with limitations of the findings. A practical interpretation of the significance of the effects is also provided in chapter 4. The criteria used for interpreting effect sizes in the meta-analysis were established by Jacob Cohen (1988). According to Cohen,

for standardized mean difference effect sizes like those reported in this meta-analysis the following criteria can be applied:

- Effect sizes ranging from 0 to .20 can be classified as trivial
- Effect sizes ranging from .20 to .50 can be classified as small
- Effect sizes ranging from .50 to .80 can be classified as large
- Effect sizes ranging from .80 and up can be classified as very large

Attenuation

Hunter and Schmidt (2015) state that:

> The goal of a meta-analysis of correlations is a description of the distribution of actual (i.e., construct-level) correlations between a given independent and a given dependent variable. If all studies were conducted perfectly, then the distribution of study correlations could be used directly to estimate the distribution of actual correlations (p. 39).

Of course, the issue is that no study is perfectly conducted, and as a result, the relationship between actual correlations and study correlations is much more complex. Hunter and Schmidt refer to study imperfections as "artifacts". They use this term as a reminder that the imperfections within a study are manmade. They outline the following eleven artifacts that can alter the value of outcome measures:

1. Sampling error
2. Error of measurement in the dependent variable
3. Error of measurement in the independent variable
4. Dichotomization of a continuous dependent variable
5. Dichotomization of a continuous independent variable
6. Range variation in the independent variable
7. Attrition artifacts: Range variation in the dependent variable

8. Deviation from perfect construct validity in the independent variable
9. Deviation from perfect construct validity in the dependent variable
10. Reporting or a transcriptional error
11. Variance due to extraneous factors that affect the relationship

Hunter and Schmidt (2015) outline a process for addressing these artifacts and the impact they have on existing data. They refer to this as the attenuation process. It is not always possible to correct for attenuation. Many times, primary studies do not include the necessary data (information needed to address the eleven artifacts) to perform these corrections. For this dissertation, corrections were made when the necessary artifact level data was available.

Appendix E has been dedicated to outlining the artifacts that have influenced the data within this dissertation. One artifact was adjusted, the reliability of the dependent measure from the studies. For this dissertation, that was the reliability of the assessment used to determine student achievement. The full list of adjustments can be found in the appendix.

Presenting the results

Results are presented in chapter 4, but methodologically, it is important to highlight that all search terms used are documented in Appendix B. The date range that searches were performed within are included with each search term in Appendix B. The number of studies that are generated using all search terms in each category are also reported in Table 3.2. All studies included are documented in Appendix D. The coding scheme used is documented in Appendix B. All data included in the meta-analysis were included in the coding scheme in Appendix B. All of this is important so the research synthesis can be checked and more importantly, so others

can continue to add to this synthesis in the future using different search terms as they continue to morph, as well as finding additional studies from other databases.

Fixed-Effect and Random-Effect Models

There are two models for meta-analysis, the fixed effect model and the random-effects model. Different assumptions are made under each model. When using the fixed-effect model, it is assumed that there is one true effect size for all of the studies in the analysis and all differences are due to sampling error. The effect size generated under a fixed-effect meta-analysis is the estimate of the common effect size (Borenstein et al., 2009). With the random-effects model, it is assumed that the true effect could be different from study to study. The summary effect in the random-effects model represents the average of the random sample of effect sizes that were generated by the studies (Borenstein et al., 2009). According to Borenstein and colleagues (2009)

> The selection of a computational model should be based on our expectation about whether or not the studies share a common effect size and on our goals in performing the analysis. It makes sense to use the fixed-effect model if two conditions are met. First, we believe that all the studies included in the analysis are functionally identical. Second, our goal is to compute the common effect size for the identified population, and not to generalize to other populations…when the researcher is accumulating data from a series of studies that had been performed by researchers operating independently, it would be unlikely that all the studies were functionally equivalent. Typically, the subjects or interventions in these studies would have differed in ways that would have impacted on the results, and therefore we should not assume a common effect size. Therefore, in these cases the random-effects model is more easily justified than the fixed-effect model (pp. 83-84).

Since this meta-analysis includes data from studies performed by separate researchers operating independently, and the studies likely differed in ways that impacted the results, for example: the subject area studied, the number of years the study was conducted, and the

multiple school levels studies, the random-effects model was more appropriate and was used for this meta-analysis. The fixed-effect results are shared for comparison purposes in addition to the random-effects model for the overall mean effect size portion of the results section only.

Cautions for a Meta-Analytic Research Method

There have been some differing views of meta-analysis in the research community. Meta-analysis has not always been widely used for educational research, and there has been some controversy around the method. One well-known controversy stems from Hanushek's meta-analysis of school expenditures and performance (Hanushek, 1989). Hanushek concluded that, "Detailed research spanning two decades and observing performance in many different educational settings provides strong and consistent evidence that expenditures are not systematically related to student achievement" (p. 49). After the report was published, there was a rather interesting back and forth between Hanushek and other researchers. Specifically, Hedges, Green, and Greenwald replicated Hanushek's original study (Hedges et. al., 1994).

Hanushek originally used a meta-analytic procedure called vote counting to perform his synthesis. Vote counting is simply comparing the number of statistically significant studies to the number of statistically nonsignificant studies to see if there is an effect (Borenstein et al., 2009). Borenstein and colleagues (2009) state that, "a nonsignificant finding could be due to the fact that the true effect is nil, it can also be due simply to low statistical power" (p. 252). They went on to argue that the process of vote counting in general is wrong. Since Hanushek (1989) used the process of vote counting and concluded that the true effect of expenditures on student achievement was nil, he made a pretty big meta-analytic mistake.

This led to differing results from the more sophisticated, combined significance tests, and combined estimation methods that Hedges and colleagues used. They concluded that there was

"strong support for at least some positive effects of resource inputs and little support for the existence of negative effects" (Hedges et al, 1994, p. 13). Hanushek then countered with an article titled, *An Exchange: Part II: Money Might Matter Somewhere: A Response to Hedges, Laine, and Greenwald* (1994). Hedges and colleagues responded again with their final article on the topic, *Money Does Matter Somewhere: A Reply to Hanushek* (1994). The titles of both of these articles sum up the nature of the exchange between the researchers. This scholarly exchange highlights the importance of considering options and meta-analytic methods before performing a research synthesis/meta-analysis. Additionally, it highlights the importance of using best practice in the method.

Chapter 4

Findings

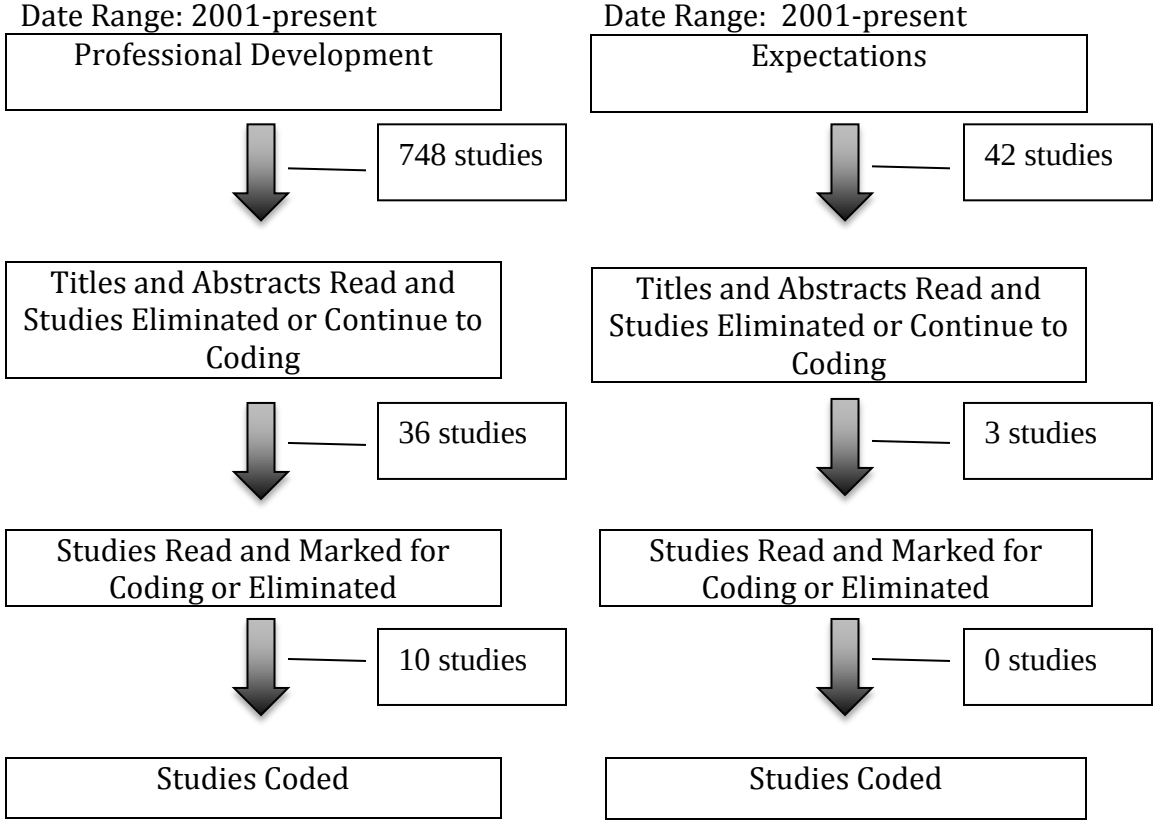
Chapter 4 begins with a review of the process implemented to identify eligible studies for this meta-analysis. A basic description of pertinent information from each study is then provided, followed by the effect size data from each study. Results are then presented including the effect sizes produced by the meta-analysis, corrections for attenuation, measures of heterogeneity, and publication bias. Moderator variables are then identified and tested by using subgroup analysis and meta-regression.

No studies regarding teacher expectations of student achievement met the eligibility criteria for this study. Ten studies regarding staff professional development between 2001 and 2014 met the eligibility criteria for inclusion in this study. Data collection was completed on October 12, 2014. Table 4.1 illustrates the review process beginning with all of the studies included using all of the search terms from Table 3.2. The two most common reasons why many of the 790 studies were excluded for this meta-analysis were:

1) They did not measure student achievement in relationship to the dependent variable (professional development or expectations).

2) They were not school-wide in nature.

Table 4.1 Flowchart of the Search Process using the ERIC database



Description of Coded Studies

The 10 studies that met the eligibility criteria for school-wide professional development were very diverse in focus. Six of the studies were commissioned by the United States Department of Education Institute for Education Sciences (I.E.S.) and were conducted by Regional Education Laboratories and/or I.E.S. The other four studies were published in highly reputable education journals; two in the American Educational Research Journal, one in the Elementary School Journal, and one in the Journal for Research in Mathematics Education. The following are the studies and a description of each in the order that they were found and coded:

1) *An Experimental Study of the Project CRISS Reading Program on Grade 9 Reading Achievement in Rural High Schools* (Kushman, Hanita, & Raphael, 2011). This study consisted of two years of professional development for staff in a program called Creating Independence

through Student-owned Strategies (CRISS). The goal of the program was to help high school students develop and use reading, comprehension, and writing strategies. This was a cluster randomized study that included 49 schools and 4,959 students. There were three effect sizes reported based upon pretest and posttest scores on the Stanford Diagnostic Reading Test. Since all three effect-sizes were based upon the same student population of ninth grade students, they were averaged for inclusion in this meta-analysis. This is an acceptable approach of combining multiple measures of the same construct from a primary study in meta-analysis (Lipsey and Wilson, 2001).

2) *Classroom Assessment for Student Learning: Impact on Elementary School Mathematics in the Central Region* (Randel, Beesley, Apthorp, Clark, Wang, Ciechinelli, & Williams, 2011). This study consisted of one year of professional development for teachers in the Classroom Assessment for Student Learning (CASL) program published by the Assessment Training Institute of Pearson Education. The training was intended to increase staff knowledge in assessment practices. The focus of the study was on measuring the impact that CASL had on student achievement in mathematics. The study was a randomized controlled design that included 67 schools and 9,596 students. One effect size was reported based upon the mathematics portion of the Colorado Assessment of Student Progress.

3) *Evaluation of Quality Teaching for English Learners (QTEL) Professional Development* (Bos, Sanchez, Tseng, Rayyes, Ortiz, & Sinicrope, 2012). This study consisted of one year of professional development in the Quality Teaching for English Learners (QTEL) developed by WestEd. The program focuses on improving teaching of secondary level English Learners. This was a randomized controlled study that included 52 schools and 36,017 students. Six effect sizes pertaining to student achievement on the California Standards Test of English

Language Arts were reported. Only the two including the whole 7th and 8th grade cohorts were used (one for the 7th grade cohort and one for the 8tth grade cohort) for this analysis since the others were subsets of the larger groups.

4) *Assessing the Value-Added Effects of Literacy Collaborative Professional Development on Student Learning* (Biancarosa, Bryk, & Dexter, 2010). This study focused on providing job embedded professional development in reading through a school-based literacy coach. The professional development outlined for this research was three years in duration. This was an accelerated multicohort, longitudinal, quasi-experimental design that included 17 schools and 8,576 students. Three effect sizes relating to student achievement were reported based upon Rasch modeling combining Dynamic Indicators of Basic Early Literacy and Terra Nova scores, one for each year of implementation. One of the three effect sizes was used for this analysis because it was gathered at the end of the professional development. The other two effect sizes were gathered after year one and year two and do not represent a true picture of the effects of the professional development over the three year timeframe.

5) *Reading First Impact Study Final Report* (Gamse, Jacob, Horst, Boulay, Unlu, Bozzi, Caswell, Rodger, Smith, Brigham, & Rosenblum, 2008). This study investigated the impact that Reading First, a one billion dollar federal initiative established by the No Child Left Behind Act aimed at ensuring that all students read at grade level, had on student reading achievement. While there were many components to Reading First, professional development was a big part of the program. Because of this, the study met the eligibility criteria for this synthesis. The study utilized a regression discontinuity design and included 248 schools and 53,271 students. Four effect sizes were reported, three that were based upon student achievement on the Stanford Achievement Test 10th Edition for Reading (one for grade 1, one for grade 2 and one for grade 3)

and one based upon the Test of Silent Word Fluency (grade 1). The two first grade effect sizes were averaged for use in this analysis. The one second grade and one third grade effect size were used as reported.

6) *Increasing Achievement by Focusing Grade-Level Teams on Improving Classroom Learning: A Prospective Quasi-Experimental Study of Title 1 Schools* (Saunders, Goldenberg, & Gallimore, 2009). The focus of this study was on measuring student achievement effects in relationship to increased school and team capacity in the area of improving student learning. The professional development for this study lasted five years, consisting of two years of professional development for the school principal followed by professional development for the principal and teacher leaders within the building. This was a quasi-experimental study that included 15 schools and 13,937 students. Ten effect sizes were reported based upon results from the Stanford Achievement Test 9th Edition given to second through fifth graders in language arts and mathematics. Only the two effect sizes from the final year gathered were used for this meta-analysis. They were averaged.

7) *Middle School Mathematics Professional Development Impact Study* (Garet, Wayne, Stancavage, Taylor, Walters, Song, Brown, Hurlburt, Zhu, Sepanik, Doolittle, & Warner, 2010). The focus of this study was on measuring the impact that one year of mathematics professional development for staff had in the area of teaching rational number concepts. This was a randomized controlled study that included 77 schools and 4,528 students. Three effect sizes were reported based upon results from the Northwest Evaluation Association (NWEA) test given to seventh graders in mathematics. The three effect sizes were averaged for use in this meta-analysis.

8) *Evaluation of the Content Literacy Continuum: Report on Program Impacts, Program Fidelity, and Contrast* (Corrin, Lindsay, Somers, Myers, Meyers, Condon, Smith, & Garcia, 2012). This study focused on the effects that the Content Literacy Continuum (CLC) intervention had on high school students' reading levels. It consisted of two years of professional development for staff. This was a randomized controlled study that included 28 schools and 9,332 students. The study reported 4 effect sizes (three effect sizes for ninth graders and one for tenth graders) based upon the results of the Group Reading Assessment and Diagnostic Evaluation. The three effect sizes for ninth graders were averaged for this meta-analysis, and the tenth grade effect size was used as reported.

9) *The Implementation and Impact of Evidence-Based Mathematics Reforms in High-Poverty Middle Schools: A Multi-Site, Multi-Year Study* (Balfanz, Mac Iver, & Byrnes, 2006). This study encompassed 4 years of professional development and implementation of the Talent Development Middle School Model mathematics Program. This was a quasi-experimental study that included 6 schools and 16,637 students. Five effect sizes were reported based upon results from the Stanford Achievement Test 9th Edition. The effect size encompassing cohort growth from the end of fifth grade to the end of eighth grade was used for this meta-analysis since students had matriculated through middle school and fully benefitted from the mathematics program at that time.

10) *Integration of Technology, Curriculum, and Professional Development for Advancing Middle School Mathematics: Three Large-Scale Studies* (Roschelle, Shechtman, Tatar, Hegedus, Hopkins, Empson, Knudsen, & Gallagher, 2009). This paper presented two randomized controlled experiments and one quasi-experiment evaluating the SimCalc approach in middle school mathematics. The study consisted of two years of professional development. It

included 73 schools and 2,984 students. Three effect sizes were reported (two for seventh grade and one for eighth grade) based upon a pretest and posttest designed by the research team. The two seventh grade effect sizes were averaged for use in this meta-analysis.

The ten studies used produced fifteen separate effect sizes that were included in this meta-analysis. All of the studies and effect sizes used, number of schools included, and number of students included can be found in table 4.2. For studies that included effect size data over multiple years of implementation the last effect size(s) reported were used in this analysis. This was done because much more professional development was delivered by the end of each multi-year study than was delivered at the end of each year. This represents a truer effect of the intended professional development on student achievement since all of the professional development was delivered by the end of the last year of each study. In addition, this approach had the advantage of enabling the analyst to investigate differences in professional development delivered over a short period of time compared to professional development delivered over many years.

This meta-analysis encompassed 632 schools and about 157,000 students. Six of the studies targeted student achievement in English/language arts and five of the studies targeted student achievement in mathematics (one study targeted achievement in English/language arts and mathematics).

Table 4.2 Study Information Included in the Meta-Analysis

Study Name	Effect Size(s) Used in Meta-analysis	Number of Schools Included in Study	Number of Students Included in Study
1) An Experimental Study of the Project CRISS Reading Program on Grade 9 Reading Achievement in Rural High Schools (2011)	-0.004	49	4,959
2) Classroom Assessment for Student Learning: Impact on Elementary School mathematics in the Central Region (2011)	-0.025	67	9,596
3) Evaluation of Quality Teaching for English Learners (QTEL) Professional Development (2012)	-0.01 0.01	52	17,837 18,180
4) Assessing the Value-Added Effects of Literacy Collaborative Professional Development on Student Learning (2010)	0.43	17	8,576
5) Reading First Impact Study (2008)	0.135 0.04 0.01	248	17,757 17,757 17,757
6) Increasing Achievement by Focusing Grade-Level Teams on Improving Classroom Learning: A Prospective, Quasi-Experimental Study of Title I Schools (2009)	.94775	15	13,937
7) Middle School mathematics Professional Development Impact Study (2010)	0.03333	77	4,528
8) Evaluation of the Content Literacy Continuum: Report on Program Impacts, Program Fidelity, and Contrast (2012)	0.12 0.10	28	4786 4546
9) The Implementation and Impact of Evidence-Based mathematics Reforms in High-Poverty Middle Schools: A Multi-Site, Multi-Year Study (2006)	0.24	6	13,637
10) Integration of Technology, Curriculum, and Professional Development for Advancing Middle School mathematics: Three Large-Scale Studies (2009)	0.565 0.56	73	2159 825
Total		632	156,837

Missing Data

None of the studies contained missing data that impacted the calculations of effect sizes in this meta-analysis. All studies reported effect size information, and were therefore included in the analysis. No provisions were taken for missing data.

Results

Research Question 1: What is the impact that creating high and appropriate expectations for all has on student achievement based on the literature?

Unfortunately, there were no studies that met the eligibility criteria for this meta-analysis in the area of staff or student expectations. Almost all of the studies that were identified in this area based upon the search terms provided by the Regional Education Lab Midwest were eliminated for the aforementioned two main reasons: 1) They did not measure an impact on student achievement, and 2) They were not school-wide in nature.

Research Question 2: What is the impact that developing staff skills at the school site have on student achievement based on the literature?

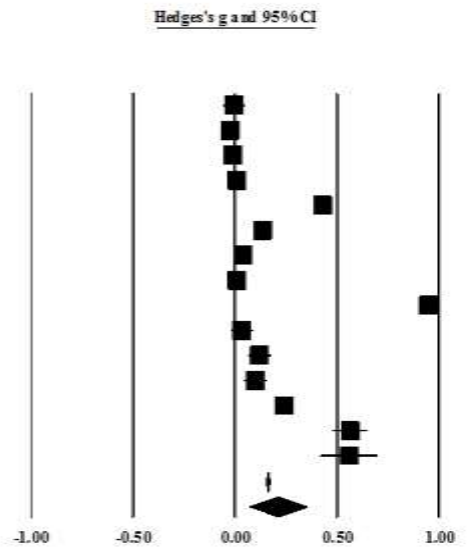
Professional development has an impact on student achievement of g= .208, which is considered a small effect. The effect sizes from the professional development studies ranged in magnitude from -.025 (small) to .948 (large). Although the random-effects model is more appropriate for this meta-analysis, both the fixed and random results are shared for the overall mean effect sizes in this section. This is done to show the reader the differences between the two models.

Throughout the rest of the dissertation, the results from the random-effect model only are used. The final effect size produced from the fixed-effects model was g=.157 (SE= .005 , Z= 30.839 , P< 0.001). The final effect size produced from the random-effects model was g=.208 (SE= .072). The Z value, which is the test of the null hypothesis that the effect size is actually 0, produced a result of 2.878 ($p < .005$). This confirms that the true mean effect size is probably not zero. This effect size falls into the small range according to the criteria established by Cohen (1988). The 95% confidence intervals were .147-.167 for the fixed-effects model and .066-.350

for the random-effects model. Figure 4.1 includes the pertinent statistics for both models.

Figure 4.1: School-wide Effects of Professional Development on Student Achievement

Model	Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Kushman et al	-0.004	0.028	0.001	-0.060	0.052	-0.141	0.888
	Randel et al	-0.025	0.020	0.000	-0.065	0.015	-1.221	0.222
	Bos et al 1	-0.010	0.015	0.000	-0.040	0.020	-0.660	0.509
	Bos et al 2	0.010	0.015	0.000	-0.019	0.039	0.673	0.501
	Biancarosa et al	0.430	0.022	0.000	0.387	0.473	19.684	0.000
	Gamse et al 1	0.135	0.015	0.000	0.106	0.164	8.985	0.000
	Gamse et al 2	0.040	0.015	0.000	0.011	0.069	2.665	0.008
	Gamse et al 3	0.010	0.015	0.000	-0.019	0.039	0.666	0.505
	Saunders et al	0.948	0.018	0.000	0.912	0.983	52.512	0.000
	Garet et al	0.033	0.030	0.001	-0.025	0.092	1.121	0.262
	Corrin et al 1	0.120	0.030	0.001	0.062	0.178	4.064	0.000
	Corrin et al 2	0.100	0.031	0.001	0.039	0.161	3.235	0.001
	Belfanz et al	0.240	0.017	0.000	0.206	0.274	13.964	0.000
	Roschelle et al 1	0.565	0.045	0.002	0.476	0.653	12.523	0.000
	Roschelle et al 2	0.559	0.073	0.005	0.416	0.703	7.616	0.000
Fixed		0.157	0.005	0.000	0.147	0.167	30.839	0.000
Random		0.208	0.072	0.005	0.066	0.350	2.878	0.004



Correction for Attenuation

Corrections for attenuation were performed based upon the reliability of the dependent variable. In this meta-analysis the dependent variable was the achievement scores on assessments identified in each study. The assessment reliability was reported or found for five of the studies. For the other five studies, the reliability was assumed to be .85. This was done because the reliability of many state assessments fall in the range of .85 (Somers, Zhu, and Wong, 2011). For a list of the corrected effect sizes used, please see Appendix E.

The corrected effect size produced from the random-effects model was g=.245 (SE= .084). The Z value, which is the test of the null hypothesis that the effect size is actually 0, produced a result of 2.915 ($p < .005$). This confirms that the true mean effect size is probably not zero in the corrected model as well. The effect size corrected for attenuation still falls into

the small range according to the criteria established by Cohen (1988). The 95% confidence intervals were .080-.410 for the random-effects model.

Measures of Heterogeneity

The test of homogeneity (Q) was significant at $p < .001$ ($Q = 2722.69$, $df = 14$), thus it was concluded that the studies included in this meta-analysis do not include a common effect size. In other words, the effect sizes vary from study to study based upon real differences in the effect size and not just because of error. This is always the case when Q is greater than the degrees of freedom. In this analysis, Q was much larger than the degrees of freedom value of 14 ($df = k-1$, where k= the number of studies).

In a meta-analysis, tau-squared is defined as, "the variance of the true effect sizes" (Borenstein et al., 2009, p. 114). Since there is also sampling error within each study, tau-squared actually represents the variance of observed effects. For this meta-analysis, tau-squared= .078. This was important to the random effects model because the amount of variance in the observed effects (tau-squared) is what determined the weights assigned for each study. Table 4.3 shows the weights assigned to each study under the fixed and random-effects models.

Table 4.3 Study Weights in Each Model

Study Number and Author	Fixed Effect Model Weight	Random Effects Model Weight
1) Kushman et al	3.23	6.67
2) Randel et al	6.21	6.71
3) Bos et al effect size 1	11.37	6.72
3) Bos et al effect size 2	11.81	6.72
4) Biancarosa et al	5.46	6.70
5) Gamse et al effect size 1	11.54	6.72
5) Gamse et al effect size 2	11.56	6.72
5) Gamse et al effect size 3	11.57	6.72
6) Sauders et al	8.0	6.71
7) Garet et al	2.95	6.67
8) Corrin et al effect size 1	2.99	6.67
8) Corrin et al effect size 2	2.73	6.66
9) Belfanz et al	8.82	6.72
10) Roschelle et al effect size 1	1.28	6.57
10) Roschelle et al effect size 2	0.48	6.30

The I^2 statistic is a descriptive statistic that "reflects the extent of overlap of confidence intervals" (Borenstein et al., 2009, pg. 118). I^2 has a range of 0-100% and allows meta-analysts to "discuss the amount of variance on a relative scale" (pg. 120). For this meta-analysis, I^2 = 99.486, which indicates almost all of the observed variance (99.486%) reflects real differences in effect sizes. This means that there was a high amount of variance between the studies that was not due to sampling error. This can be explained by the study level covariates, also known as moderator variables. Further details are provided in the moderator variable section of this chapter.

Publication Bias

Several statistical tests were run to check for publication bias. It was important to check for publication bias as it is common that published studies are more likely to report larger effect sizes than studies that are not published (Borenstein et al., 2009). The first test was the Fail Safe

N, which reports, "how many new studies averaging a null result are required to bring the overall treatment effect to nonsignificance" (Cooper et al., 2009, p. 442). For the present meta-analysis, the Fail Safe N= 4,098. This means that there would need to be 4,098 missing studies that had zero effect to make the results non-significant.

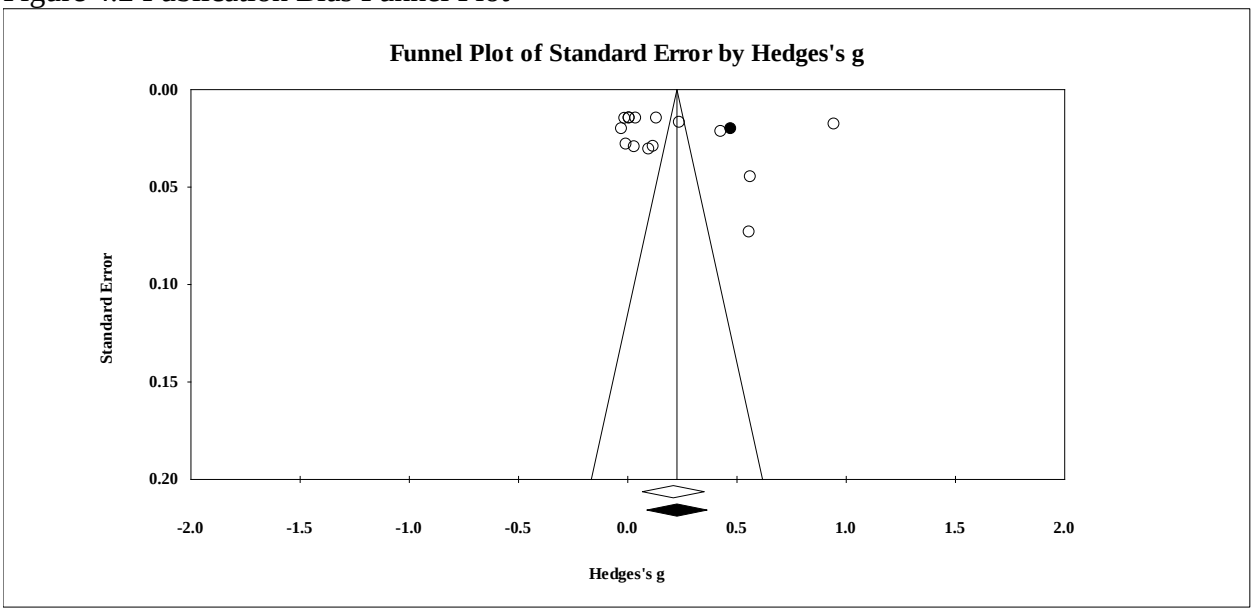
Another step in checking for publication bias is to produce a standard error funnel plot by running Duval and Tweedie's trim and fill procedure (Borenstein et al., 2009). Figure 4.2 represents the standard error funnel plot for this meta-analysis. Typically, larger, more precise studies cluster at the top of the funnel plot near the mean effect size. Smaller studies are generally more spread out toward the bottom of the plot. If publication bias exists, the funnel plot is likely to be asymmetrical (Borenstein et al., 2009). Typically, asymmetrical funnel plots are due to smaller missing studies (studies that were not published due to the aforementioned publication bias). When this occurs, the lack of symmetry is visible on the bottom left-hand side of the funnel plot.

Figure 4.2 represents the standard error funnel plot for this meta-analysis with the trim and fill results added. There are a few patterns that emerge from this funnel plot. First, there are not many small scale studies included. This is evidenced by the low standard error from the studies (they are mostly clustered near the top of the plot). Second, most of the effect sizes are clustered to the left of the mean effect size. This indicates that it is not likely that studies are missing due to publication bias. Third, when a trim and fill analysis was performed to search for missing effect size that would change the mean effect size, it was determined that one study was missing with a larger effect size than the mean effect size. The filled or imputed study is represented by the black dot on the plot. The open circles represent the original studies. Missing studies with higher effect sizes than the mean effect size are unusual. The trim and fill changed

the overall mean effect size for the random-effects model from g=.208 (95% CI= .066-.350) to g=.225 (95% CI=.087-.363). This still falls in the small effect range. The difference can be seen in Figure 4.2. The original random-effects results are represented by the open diamond. The trim and fill results are represented by the black (filled in) diamond.

In the present meta-analysis there do not appear to be missing studies with lower effect sizes. In addition, there appears to be a downward bias in the summary effect, therefore, publication bias does not appear to exist. In addition, there would need to be 4,098 missing studies that had zero effect to make the results non-significant. Even with the downward bias in the summary effect, the results of this analysis are still very similar when the bias was adjusted for.

Figure 4.2 Publication Bias Funnel Plot



Moderator Variables

While this meta-analysis focused on the impact that professional development had on student achievement, other variables often moderate or mediate the relationship between two

variables in a meta-analysis (Cooper et al, 2009). These additional variables are called moderator variables. This impact can be analyzed using the Comprehensive Meta-Analysis 2 Software by running subgroup analysis for categorical moderator variables and meta-regression for continuous moderator variables. For this meta-analysis, the moderator variables were identified as:

1) Publication type: which was a categorical variable where studies were classified as and Institute of Education Sciences Publication or an Academic Journal Publication.

2) Subject: which was a categorical variable where studies were classified as focusing on language arts or mathematics.

3) School type: which was a categorical variable where studies were classified as elementary or secondary.

4) Amount of professional development delivered: which was considered as both a categorical variable and a continuous variable. The categorical variable classified studies as two years or less of professional development provided, or two years or more professional development. The continuous variable listed studies based upon the number of years that professional development was delivered.

It is important to note that study number six reported results based upon an overall composite score obtained from the Stanford Achievement Test 9th Edition that included language arts and mathematics measures. Since more of the subtests were language arts in nature, this study was grouped with the language arts group in the subgroup analysis. It is also noteworthy that the school type was based upon what the schools were reported as in the primary study. Middle schools and high schools were grouped together and are referred to as Secondary Schools. A list of grade levels included in each study can be found in the final coding scheme in

Appendix B. Finally, it would be ideal to have the total number of hours of provided professional development for staff in each study. This would be a more conducive way to consider differences in professional development based upon a continuous time variable since the amount of professional development delivered per year could vary. Since this could not be derived from all of the studies, it was decided to break the studies into groups based upon the number of years that professional development was delivered.

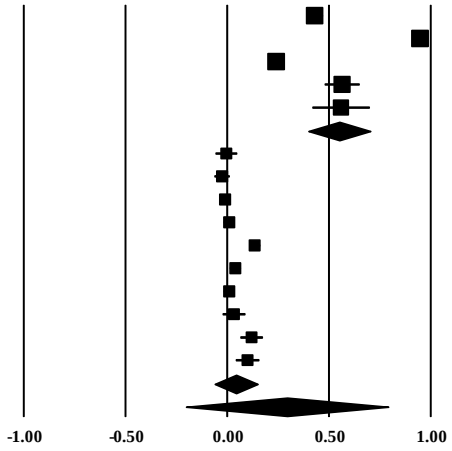
Since real differences in effect sizes between studies exist in this meta-analysis, as revealed in the various tests of heterogeneity performed, it is important to perform moderator analysis to attempt to explain why. Comparing the mean effect of different subgroups in a meta-analysis is like analyzing variance in a primary study (Borenstein et al., 2009). The analysis of these moderator variables was done based upon the random-effects model.

The first moderator variable investigated was that of publication type. Four of the studies and five of the effect sizes came from studies published by academic journals. Six of the studies and ten of the effect sizes used in this meta-analysis came from studies published by the I.E.S. When these two groups were compared, there were big differences in the effect sizes between the groups. The results can be viewed in figure 4.3. The mean effect size generated for the studies from academic journals is g= .548 (SE= .078, Z= 7.056, $p < .001$). The 95% confidence interval runs from .396 to .700. The mean effect size from the I.E.S. studies is g= .041 (SE= .054). The 95% confidence interval runs from -.065 to .146. The test that the mean is not zero is addressed by the Z value and corresponding p-value (Z= .756, p= .450). This does not meet the criterion of 0.05. Thus, we cannot reject the null hypothesis. This means we cannot conclude that the effect size of I.E.S. studies is not zero.

This represents a large difference in the mean effect size of at least .507. In addition, the difference between studies published by I.E.S. and those published by professional journals was statistically significant (Q* between= 28.78, df= 1, p < .001). In other words, the effect size probably does differ based upon the publication type. This certainly helps explain the variance that was identified within this meta-analysis by the tests of heterogeneity that were performed. There are real differences in the effect sizes between studies within this meta-analysis, and one of the reasons is the differences between I.E.S. studies and studies published in academic journals.

Figure 4.3 Effects of Professional Development by Publication Type

Group by Publication Type	Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Article	Biancarosa et al	0.430	0.022	0.000	0.387	0.473	19.684	0.000
Article	Saunders et al	0.948	0.018	0.000	0.912	0.983	52.512	0.000
Article	Belfanz et al	0.240	0.017	0.000	0.206	0.274	13.964	0.000
Article	Roschelle et al 1	0.565	0.045	0.002	0.476	0.653	12.523	0.000
Article	Roschelle et al 2	0.559	0.073	0.005	0.416	0.703	7.616	0.000
Article		0.548	0.078	0.006	0.396	0.700	7.056	0.000
IES Study	Kushman et al	-0.004	0.028	0.001	-0.060	0.052	-0.141	0.888
IES Study	Randel et al	-0.025	0.020	0.000	-0.065	0.015	-1.221	0.222
IES Study	Bos et al 1	-0.010	0.015	0.000	-0.040	0.020	-0.660	0.509
IES Study	Bos et al 2	0.010	0.015	0.000	-0.019	0.039	0.673	0.501
IES Study	Gamse et al 1	0.135	0.015	0.000	0.106	0.164	8.985	0.000
IES Study	Gamse et al 2	0.040	0.015	0.000	0.011	0.069	2.665	0.008
IES Study	Gamse et al 3	0.010	0.015	0.000	-0.019	0.039	0.666	0.505
IES Study	Garet et al	0.033	0.030	0.001	-0.025	0.092	1.121	0.262
IES Study	Corrin et al 1	0.120	0.030	0.001	0.062	0.178	4.064	0.000
IES Study	Corrin et al 2	0.100	0.031	0.001	0.039	0.161	3.235	0.001
IES Study		0.041	0.054	0.003	-0.065	0.146	0.756	0.450
Overall		0.291	0.254	0.064	-0.206	0.788	1.149	0.251



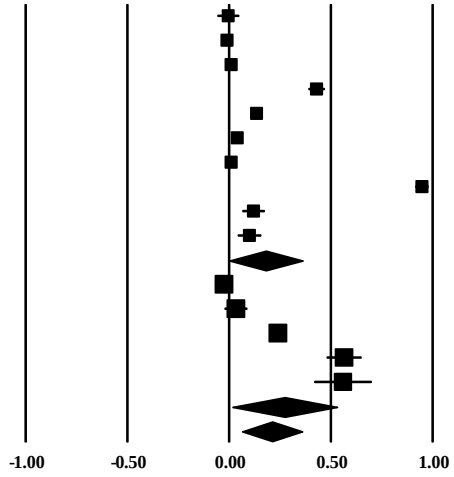
The second moderator was subject area. Six of the studies and ten effect sizes included in the analysis were based on language arts results. Four of the studies and five effect sizes were based upon mathematics results. It cannot be concluded that subject area has an impact on the overall mean effect size in this meta-analysis. The results of the subgroup analysis can be

viewed in figure 4.4. The mean effect size generated for the studies from language arts is g= .178 (SE= .092). The test that the mean is not zero is addressed by the Z value and corresponding p-value (Z= 1.933, p= .053). This does not meet the criterion of 0.05. Thus, the null hypothesis that the mean effect for language arts is zero cannot be rejected. The 95% confidence interval runs from -.003 to .358. The mean effect size from mathematics studies is g= .270 (SE= .131, Z= .2.061, p = .039). For the mathematics effect, it is concluded that the mean effect size is probably not zero. The 95% confidence interval runs from .013 to .528.

In addition, the Q test between language arts and mathematics (Q* between= .333, df= 1, p = .564) revealed that there are not real differences in effect size based upon subject. As a result, this moderator does not help explain the variance in the overall meta-analysis.

Figure 4.4 Effects of Professional Development by Subject Area

Group by Subject Area	Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Language Arts	Kushman et al	-0.004	0.028	0.001	-0.060	0.052	-0.141	0.888
Language Arts	Bos et al 1	-0.010	0.015	0.000	-0.040	0.020	-0.660	0.509
Language Arts	Bos et al 2	0.010	0.015	0.000	-0.019	0.039	0.673	0.501
Language Arts	Biancarosa et al	0.430	0.022	0.000	0.387	0.473	19.684	0.000
Language Arts	Gamse et al 1	0.135	0.015	0.000	0.106	0.164	8.985	0.000
Language Arts	Gamse et al 2	0.040	0.015	0.000	0.011	0.069	2.665	0.008
Language Arts	Gamse et al 3	0.010	0.015	0.000	-0.019	0.039	0.666	0.505
Language Arts	Saunders et al	0.948	0.018	0.000	0.912	0.983	52.512	0.000
Language Arts	Corrin et al 1	0.120	0.030	0.001	0.062	0.178	4.064	0.000
Language Arts	Corrin et al 2	0.100	0.031	0.001	0.039	0.161	3.235	0.001
Language Arts		0.178	0.092	0.008	-0.003	0.358	1.933	0.053
Math	Randel et al	-0.025	0.020	0.000	-0.065	0.015	-1.221	0.222
Math	Garet et al	0.033	0.030	0.001	-0.025	0.092	1.121	0.262
Math	Belfanz et al	0.240	0.017	0.000	0.206	0.274	13.964	0.000
Math	Roschelle et al 1	0.565	0.045	0.002	0.476	0.653	12.523	0.000
Math	Roschelle et al 2	0.559	0.073	0.005	0.416	0.703	7.616	0.000
Math		0.270	0.131	0.017	0.013	0.528	2.061	0.039
Overall		0.208	0.075	0.006	0.061	0.356	2.766	0.006

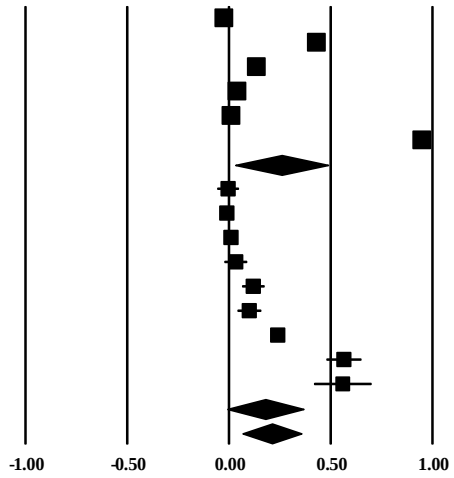


The third moderator was school type. Four of the studies and six of the effect sizes in this meta-analysis came from elementary schools. Six of the studies and nine of the effect sizes came from secondary schools.[14] The results can be viewed in figure 4.5. The mean effect size generated for the studies from Elementary Schools is g= .256 (SE= .116, Z= 2.203, p = .028). The null hypothesis that the effect size is zero can be rejected. The 95% confidence interval runs from .028 to .484. The mean effect size from Secondary School is g= .176 (SE= .095, Z= 1.845, p = .065). The null hypothesis that the effect size is zero cannot be rejected with Secondary Schools. The Q test between the two school types indicated that there is not an association between the school type and the total mean effect in this study (Q* between= .283, df= 1, p = .595). This moderator does not help to explain any of the variance in this meta-analysis.

[14] The primary studies reported the schools as elementary, middle school, or high school. When the analysis was run with all three categories there were only two effect sizes for high schools and four for middle schools. As a result, they were combined for this analysis.

Figure 4.5 Effects of Professional Development by School Type

Group by School Type	Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Elementary	Randel et al	-0.025	0.020	0.000	-0.065	0.015	-1.221	0.222
Elementary	Biancarosa et al	0.430	0.022	0.000	0.387	0.473	19.684	0.000
Elementary	Gamse et al 1	0.135	0.015	0.000	0.106	0.164	8.985	0.000
Elementary	Gamse et al 2	0.040	0.015	0.000	0.011	0.069	2.665	0.008
Elementary	Gamse et al 3	0.010	0.015	0.000	-0.019	0.039	0.666	0.505
Elementary	Saunders et al	0.948	0.018	0.000	0.912	0.983	52.512	0.000
Elementary		0.256	0.116	0.014	0.028	0.484	2.203	0.028
Secondary	Kushman et al	-0.004	0.028	0.001	-0.060	0.052	-0.141	0.888
Secondary	Bos et al 1	-0.010	0.015	0.000	-0.040	0.020	-0.660	0.509
Secondary	Bos et al 2	0.010	0.015	0.000	-0.019	0.039	0.673	0.501
Secondary	Garet et al	0.033	0.030	0.001	-0.025	0.092	1.121	0.262
Secondary	Corrin et al 1	0.120	0.030	0.001	0.062	0.178	4.064	0.000
Secondary	Corrin et al 2	0.100	0.031	0.001	0.039	0.161	3.235	0.001
Secondary	Belfanz et al	0.240	0.017	0.000	0.206	0.274	13.964	0.000
Secondary	Roschelle et al 1	0.565	0.045	0.002	0.476	0.653	12.523	0.000
Secondary	Roschelle et al 2	0.559	0.073	0.005	0.416	0.703	7.616	0.000
Secondary		0.176	0.095	0.009	-0.011	0.363	1.845	0.065
Overall		0.208	0.074	0.005	0.064	0.353	2.824	0.005



Effects of the Duration of Professional Development on Student Achievement

A third research question emerged as the present meta-analysis progressed. It became clear that not all studies investigated the impact of professional development over the same time period. Thus, research question three emerged:

What is the impact that differences in the duration of professional development have on student achievement?

Analyzing the variable as a moderator provided the answer to this question. The fourth and final moderator for the present meta-analysis was the number of years professional development was delivered. This moderator was considered in two different ways:

1) First, as a categorical variable where studies were classified as two years or less of professional development or two years or more.

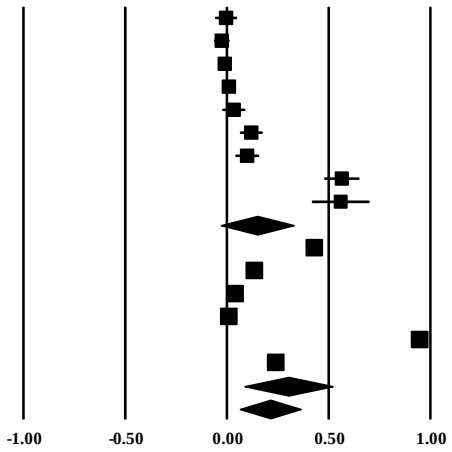
2) Second, as a continuous variable based upon the total number of years of professional

development.

A subgroup comparison was performed for the categorical variable and a meta-regression for the continuous. Six of the studies and nine of the effect sizes in this meta-analysis came from results of two years or less of professional development. Four of the studies and six of the effect sizes came from the results of more than two years of professional development. When these two groups were compared, there were moderate differences in the effect sizes between the two groups. The results can be viewed in figure 4.6. The mean effect size generated for the studies from two years or less of professional development is g=.146 (SE= .091, Z= 41.612, p = .107) The null hypothesis that the effect size is zero cannot be rejected for two years or less of professional development. The mean effect size from more than two years of professional development is g= .30 (SE= .110, Z= 2.724, p = .006). The 95% confidence interval runs from .084 to .516. The effect size is not likely zero for more than two years of professional development. In addition the Q statistic comparing two years or less of professional development and more than two years of professional development was Q*between = 1.168, df = 1, and p= .280. As such, it cannot be concluded that the amount of professional development has an impact on the overall mean effect size. It is however, noteworthy that the mean effect size for the studies that delivered more than two years of professional development generated an effect size of g = .30. This is larger than the mean effect size for the entire meta-analysis that was g= .210.

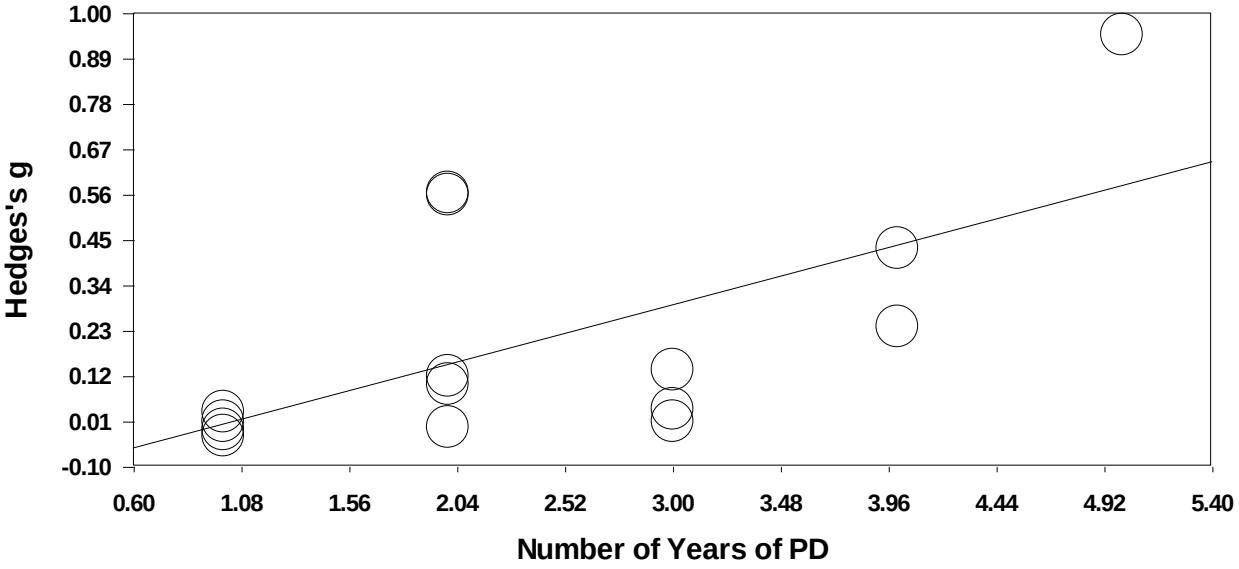
Figure 4.6 Effects of Professional Development by Number of Years PD was Provided

Group by # of Years of PD	Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
2 years or less	Kushman et al	-0.004	0.028	0.001	-0.060	0.052	-0.141	0.888
2 years or less	Randel et al	-0.025	0.020	0.000	-0.065	0.015	-1.221	0.222
2 years or less	Bos et al 1	-0.010	0.015	0.000	-0.040	0.020	-0.660	0.509
2 years or less	Bos et al 2	0.010	0.015	0.000	-0.019	0.039	0.673	0.501
2 years or less	Garet et al	0.033	0.030	0.001	-0.025	0.092	1.121	0.262
2 years or less	Corrin et al 1	0.120	0.030	0.001	0.062	0.178	4.064	0.000
2 years or less	Corrin et al 2	0.100	0.031	0.001	0.039	0.161	3.235	0.001
2 years or less	Roschelle et al 1	0.565	0.045	0.002	0.476	0.653	12.523	0.000
2 years or less	Roschelle et al 2	0.559	0.073	0.005	0.416	0.703	7.616	0.000
2 years or less		0.146	0.091	0.008	-0.032	0.324	1.612	0.107
more than 2 years	Biancarosa et al	0.430	0.022	0.000	0.387	0.473	19.684	0.000
more than 2 years	Gamse et al 1	0.135	0.015	0.000	0.106	0.164	8.985	0.000
more than 2 years	Gamse et al 2	0.040	0.015	0.000	0.011	0.069	2.665	0.008
more than 2 years	Gamse et al 3	0.010	0.015	0.000	-0.019	0.039	0.666	0.505
more than 2 years	Saunders et al	0.948	0.018	0.000	0.912	0.983	52.512	0.000
more than 2 years	Belfanz et al	0.240	0.017	0.000	0.206	0.274	13.964	0.000
more than 2 years		0.300	0.110	0.012	0.084	0.516	2.724	0.006
Overall		0.210	0.076	0.006	0.061	0.359	2.767	0.006



A meta-regression was performed as well, based upon the number of years professional development was delivered. The results can be viewed in figure 4.7. For this model, Q= 11.349 with df= 1 and $p < .001$, therefore, it can be concluded that there is a correlation between the number of years of professional development and the effect size. The slope under this random-effects meta-regression equals .145. For every additional year that professional development is provided, the model predicts an increase in effect size of .145. The 95% confidence interval for the slope runs from .060 to .229. In other words, the true coefficient for the slope could be as low as .060 or as high as .229. In this meta-regression, R^2= .50. In other words, the number of years professional development is provided helps to explain 50% of the between study variance in the overall meta-analysis.

Figure 4.7 **Regression of Number of Years of PD on Hedges's g**



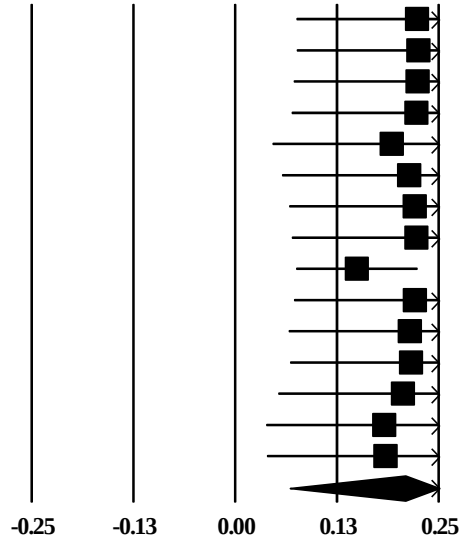
Sensitivity Analysis

When performing a meta-analysis, it is important to perform sensitivity tests to ensure no one study disproportionately affected the mean effect size in the meta-analysis. One way to do this is by performing the one study removed test using the Comprehensive Meta-Analysis Software. This test shows what the mean overall effect size would be if the study was removed. Figure 4.8 shows the results of this test. In this meta-analysis, there were no studies that disproportionately affected the mean effect size. The largest change in effect size would be .058 if the Saunders et al. study were removed. The mean effect size in the random-effects model would change from g= .208 to g= .15. Additionally, while the Saunders et al. study appears to be an outlier based upon the reported effect size of .948, this particular study reported on effect sizes for each year of professional development over a five-year sustained period. The results from years one and two, which is all a majority of the studies in the analysis provided, were more

similar to the other studies (g= .1815 and g= .2261)[15] respectively. Since the effect sizes from years one and two were very similar to the other effect sizes utilized for this meta-analysis, and the Saunders et al. study included effect sizes over a five year period it was not considered an outlier for the present meta-analysis.

Figure 4.8 Effects with Study Removed

Study name	Point	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Kushman et al	0.224	0.076	0.006	0.075	0.372	2.945	0.003
Randel et al	0.225	0.077	0.006	0.075	0.375	2.943	0.003
Bos et al 1	0.224	0.078	0.006	0.072	0.377	2.878	0.004
Bos et al 2	0.223	0.078	0.006	0.069	0.377	2.839	0.005
Biancarosa et al	0.192	0.075	0.006	0.045	0.339	2.565	0.010
Gamse et al 1	0.214	0.080	0.006	0.057	0.371	2.674	0.008
Gamse et al 2	0.221	0.079	0.006	0.066	0.375	2.793	0.005
Gamse et al 3	0.223	0.078	0.006	0.069	0.376	2.841	0.004
Saunders et al	0.150	0.038	0.001	0.074	0.225	3.900	0.000
Garet et al	0.221	0.076	0.006	0.072	0.370	2.906	0.004
Corrin et al 1	0.215	0.076	0.006	0.065	0.364	2.816	0.005
Corrin et al 2	0.216	0.076	0.006	0.067	0.365	2.840	0.005
Belfanz et al	0.206	0.078	0.006	0.052	0.360	2.627	0.009
Roschelle et al 1	0.183	0.074	0.006	0.038	0.329	2.466	0.014
Roschelle et al 2	0.185	0.075	0.006	0.039	0.331	2.476	0.013
	0.208	0.072	0.005	0.066	0.350	2.878	0.004



The fourth and final research question was:

Of the two variables studied, which has a greater impact on student achievement based on the literature?

Since there were no studies that met the eligibility criteria for the first variable, expectations of students, an analysis for this research question cannot be provided. While studies were found that reported the effects that staff expectations of students had on student

[15] Note: Two effect sizes were actually reported for each year. They were averaged.

achievement, they were not school-wide in nature and, therefore, they did not meet the eligibility criteria for this meta-analysis.

Chapter 5

Discussion

Public opinion against public schools has continually increased over the last several decades. Major events like Sputnik, the "A Nation at Risk" report, comparisons made between the United States and other industrialized nations on international examinations, and the passage of more school laws like the No Child Left Behind Act have caused the American public to believe the schools are failing to do their jobs. The response from many professional educators in schools has been to work feverishly to attempt to fulfill the new mission of schools; to ensure high levels of learning for all students. A sizeable research base exists in how to do this at the school level. Much of this work comes from the Effective Schools movement that began after the release of the Coleman Report in 1966.

The Coleman report indicated that schools had very little impact on student success or failure. The assertion was that socioeconomic factors had the largest relationship to student achievement, and that, schools only accounted for a minimal amount of the differences in student achievement (Coleman et al., 1966). While the impact that socioeconomic factors have on student achievement cannot be denied, the effective schools research has highlighted the fact that schools do in fact have an impact on student achievement as well.

Effective Schools Research Discussion

Effective schools research has gone through many phases and has been continually refined over the last several decades. The work has generated a widely agreed upon list of factors that influence student achievement in schools, also referred to as correlates. Two of these correlates were the focus of this dissertation: school-wide professional development, and expectations. While there have been several studies done identifying the correlates, to this point,

a meta-analysis of these two variables has not been conducted. This gap in the literature served as the purpose of this dissertation.

Discussion of Findings

Specifically, this meta-analysis sought to find the effects of school-wide student and staff expectations on student learning, and the effects of school-wide professional development on student learning. Comparing the effects of professional development and staff expectations on student learning was also a goal since they are part of a broader list of effective schools correlates. None of the studies on student or staff expectations met the eligibility criteria for this analysis. Therefore, no results could be generated regarding school-wide expectations. In addition, the difference in effects between school-wide expectations and school-wide professional development could not be compared.

Findings were shared for research question two which was, what is the impact that developing staff skills at the school site have on student achievement based on the literature? The random-effects mean effect size for school-wide professional development from studies published since 2001 was g= .208. When corrections for attenuation were performed the effect size was g=. 245. Both of these effects fall into the small range.

Discussion of Sensitivity Analysis, Missing Data, and Variance

The sensitivity analysis testing the impact that each study had on the overall mean effect size revealed that no single study impacted the overall mean effect size by more than .058 standard deviation units. The .058 impact came from the Saunders et al. study which implemented a five year approach to providing professional development. The study also reported effect sizes for years one through four that were very similar to the other results included in this analysis. All studies that met the inclusion criteria were left in the final meta-

analysis.

There was no missing data from any of the studies that met the inclusion criteria for this meta-analysis. All effect size data were included; therefore, all eligible studies were included in the analysis. In addition, tests of publication bias were run. It is not unusual for these tests to indicate that smaller scale studies with lower effect sizes might be missing in many meta-analyses due to fewer of such studies being published. This occurs because many smaller scale studies that do not report a large effect are not published and therefore are less available to meta-analysts. This was not the case for the present meta-analysis. In fact, the tests for publication bias indicated that there could possibly be one smaller sample size study missing with a higher effect size than the mean effect size in this meta-analysis. This potential missing study would have changed the overall mean effect size from g= .208 to g= .225. In addition, there would need to be 4,098 studies with reporting 0 effects for professional development for the effect size to become non-significant.

Most of the variance in this meta-analysis was between studies. 99.486% of the variance was between study variance, and .514% of the variance was within study variance due to sampling error. To speculate on the causes of the between study variance, moderator analysis was performed.

Discussion of Moderator Analyses

Four moderator variables were investigated in this dissertation. The first was publication type. The ten studies all came from two publication types: Institute of Education Sciences published studies and studies published by academic journals. There were significant differences in the mean effect size produced by the two publication types. The Institute of Education Sciences published studies produced an effect size of g= .041 and it could not be ruled out that

the effect was actually zero. This is in contrast to an effect size of g= .548 for academic journals. This represents a large difference in effect sizes. My conclusion is that the effects included in this study are related to the publication type.

There could be several reasons why this variance exists: First, I.E.S. studies are typically very large, comprehensive, and as a result expensive. This could lead to the studies being shorter in duration to those produced by independent researchers and published in academic journals. The I.E.S. studies included in this meta-analysis were certainly shorter in duration than those published in academic journals. The average amount of professional development delivery in the I.E.S. studies was 1.67 years. The average amount of professional development delivery in the academic journal studies was 3.75 years. As was highlighted in chapter 4, there is a correlation between the number of years professional development was delivered and the corresponding effect size.

The second moderator investigated was subject area. A link between the subject area and the overall mean effect size could not be made based upon the subgroup analysis. The third moderator was school type. This moderator did not help to explain any of the between study variance in the overall meta-analysis. The subgroup analysis revealed that a link between school type and the mean effect size could not be established.

The final moderator investigated was the number of years professional development was delivered. As shared in chapter 4, this was investigated as a categorical variable and as a continuous variable. The subgroup analysis of the categorical variable (two years or less of professional development as compared to more than two years of professional development) could not establish a link between the numbers of years professional development was delivered and the overall mean effect size in the meta-analysis.

However, the meta-regression of the continuous variable (number of years professional development was delivered) did produce a relationship between the number of years of professional development and the overall mean effect size in the meta-analysis. The slope indicates that for each additional year that professional development was delivered an increase in the effect size of .145 was generated. It makes intuitive sense that the longer a staff in a school has support in the form of professional development in a given area, the more proficient they will become in that area. This meta-analysis reveals that this also impacts student learning, and the longer the staff in the school are exposed to the professional development the greater the achievement gains for students.

Comparison of These Findings to Others Pertaining to Professional Development

This was the first meta-analysis studying the effects that school-wide professional development has on student achievement. Since many schools and school districts provide professional development that is school-wide in nature, the results of this meta-analysis are important in their own right and should be considered differently from previous meta-analyses on this topic.

While this is the first meta-analysis on the school-wide relationship between professional development and student achievement, as noted in chapter two of this dissertation, there have been several meta-analyses on professional development. The most comprehensive analysis of the impact of professional development was done by Hattie (2012). The effect size of professional development reported by Hattie was .51. While this is quite a bit higher than the .208 effect size reported in this dissertation, it is important to note that the Hattie (2012) analysis included the impact that professional development had on changes in teacher learning, teacher

behavior, teacher reactions to professional development, and teacher skills as well as changes in student achievement. This was a much different approach than was taken with this dissertation. The only factors that were the focus of this dissertation were the impact professional development had on student achievement. With this in mind, the .208 effect size reported in this dissertation does not seem out of the ordinary. In addition, since the goal of much of the professional development that is delivered in school is to increase student achievement, this study provides results that will be very useful to the field. The school-wide nature of the results of this meta-analysis makes it extremely beneficial as well since many schools and districts consistently provide this type of professional development.

Limitations

There are several limitations to this dissertation. First, most of the studies included in this dissertation were randomized controlled studies or quasi-experimental studies with control schools. In all cases, the control schools still received professional development. It was simply the professional development that the schools already had in place and were planning to deliver. These control schools simply did not receive the specific professional development that was studied at the treatment schools. This likely led the effect size of .208 that is reported in this dissertation, to be a very conservative low estimate of the actual effect that professional development has on student achievement. This is the case because all schools, even the control schools, received professional development. It would be unethical for schools to receive no professional development during the course of a study, therefore, randomized controlled studies might not be best suited for measuring the impact that school-wide professional development has on student achievement.

Another limitation of this dissertation was that the results only pertain to language arts and mathematics. Since no studies of the impact of professional development on other subjects were found, it cannot be assumed that a .208 effect size would be generated for other subject areas. In addition, this particular meta-analysis was a limited research review. The only database used was the ERIC database.

Finally, this meta-analysis was limited in scope to articles meeting the eligibility criteria from 2001 to the present. It is likely that studies exist prior to 2001 that meet the eligibility criteria.

Implications

The results of this dissertation do have implications on both policy and practice. From a policy perspective, many state and federal funds are devoted to professional development each year, including federal Title I and Title II funds. When these funds are used for building-wide professional development in language arts and mathematics, it can be expected that student achievement will increase. While the effect size of .208 is considered small, unlike many interventions, this effect is for every student who attends the school. In addition, legislators, departments of education, and those with authority over how the funds are spent should insist on longer term planning and use of professional development dollars and avoid letting schools spend funds on one-year professional development solutions switching focus from year to year. Solid evidence exists, based upon the results of this dissertation that the longer professional development is sustained the higher the effect it will generate on student achievement.

The same implication exists for practice. The average effect of .208 might seem small at first glance, but it is important to remember that this effect is for every student who attends the

school delivering the professional development. In addition, the present study also showed the longer the school can sustain the professional development focus, the larger the effect will be. Schools should develop long-term professional development plans and stick to them over time. This would increase the effect that the professional development effort has on student achievement. While it is tempting for schools to jump to the next best thing, or a new initiative each year, the results of this dissertation illuminate the fact that short-term professional development does not provide the most "bang for the buck" compared to longer term multi-year professional development. The results of this dissertation suggest that professional development plans for schools should be thought out on the front end and then implemented for multiple years. If the goal of the professional development is to increase student learning, than a multi-year approach is necessary to maximize the effect.

Future Research

Several potential areas of future research were illuminated throughout this dissertation. The most obvious is measuring the relationship between school-wide staff expectations for students and student achievement. While there have been a few studies conducted on measuring the impact that individual teacher expectations have on student achievement, no studies were found measuring expectations of the staff of an entire school. It is possible that the search terms used in this meta-analysis did not accurately capture pieces of this work. Future work could look at potential links between Roger Goddard's work around collective efficacy and school-wide staff expectations for student learning as well as potential links between school-wide academic emphasis and school-wide academic optimism. While it was not the focus of this dissertation, there could be natural links in these areas to staff expectations for student learning. Once links

are made or there are several studies done on school-wide staff expectations for student learning, a meta-analysis would be beneficial as well.

Future research on the effects that professional development has on student achievement in subjects other than language arts and mathematics would be beneficial as well. The only subject areas included in the studies in this dissertation were language arts and mathematics. As schools continue to look for ways to increase student achievement in multiple subject areas, knowing the impact that professional development has on student achievement in all subject areas would be beneficial when allocating funds and planning for long term professional development.

Finally, all of the studies focused on professional development included in the meta-analysis for this dissertation focused on specific professional development that was delivered to entire schools or grade levels. In other words, all staff in the school or grade level received the same professional development that had the same focus. In addition, most of the studies compared this approach to control schools that continued professional development as usual. It would be unethical to stop all professional development for school staff, but other quasi-experimental design studies measuring the effects of school-wide professional development that school staff were committed to might be beneficial.

When measuring the effect that professional development has on student achievement, randomized controlled studies might not be the most beneficial for a few reasons. First, randomized controlled studies are expensive to conduct. As a result, they do not typically last for multiple years. As evidenced by this dissertation, professional development that is sustained over multiple years is likely more impactful on student achievement. Second, school staff might not be committed to the professional development delivered in randomized controlled studies.

School staff who are greatly committed to a particular focus of professional development would have a hard time being a part of a control group and not implementing the professional development.

It would also be beneficial to see if school-wide professional development that is differentiated for staff based upon their needs would increase the effect on student achievement. None of the studies in this meta-analysis took that approach. One way this could be done is by having staff set growth goals and offering professional development to all staff in a school based upon the area they targeted for growth. Since none of the professional development included in the studies took this approach, it is not possible to determine if this would lead to greater improvement in student achievement. This type of research would be helpful to those working in schools.

Conclusion

The research base surrounding effective schools is already very large. This dissertation has added to that research base by identifying necessary future research in the area of school-wide expectations. In addition, professional development has been identified as one of the factors in effective schools that influence student achievement. This dissertation has identified the effect of school-wide professional development on student learning.

The findings are useful outside of effective schools research as well, as they pertain to school-based professional development in general. While the meta-analysis produced a small effect of g= .208, it is important to remember that this effect is for every student in a school. In addition, this is a very conservative estimation of the effect size since many of the control groups

in the randomized controlled studies included in this meta-analysis also provided professional development, though not the specific professional development that was being studied.

Professional development has occurred in schools for decades and continues every day across the United States. The findings of this dissertation indicate that professional development should continue in schools, and that professional development is more effective when delivered over multiple years. The meta-regression performed in this meta-analysis indicated that for each consecutive year that professional development is continued, an increase in the effect size of .145 can be expected. This is clear evidence that school-based professional development plans should last multiple years.

References

Balfanz, R., Mac Iver, D. J., & Byrnes, V. (2006). The implementation and Impact of Evidence-Based mathematics Reforms in High-Poverty Middle Schools: A Multi-Site, Multi-Year Study. *Journal for Research in mathematics Education*, 37(1), 33-64.

Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literary collaborative professional development on student learning. *The Elementary School Journal, 111(1)*, 7-34.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. United Kingdom: John Wiley & Sons, Ltd.

Bos, J., Sanchez, R., Tseng, F., Rayyes, N., Ortiz, L., and Sinicrope, C. (2012). *Evaluation of Quality Teaching for English Learners (QTEL) Professional Development* (NCEE 2012-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bosker, R.J. (1992). *The stability and consistency of school effects in primary education.* Enschede: University of Twente.

Brookover, W. B., Schweitzer, J. H., Schneider, J. M., Beady, C. H., Flood, P. K., & Wisenbaker, J. M. (1978). Elementary School Social Climate and School Achievement. American Educational Research Journal, 15(2), 301-318.

Bryk, A., & Raudenbush, S. (1988-1989). Methodological Advances in Analyzing the Effects of Schools and Classrooms on Student Learning. Review of Research in Education, 15, 423-475.

Byrk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* New York: Sage

Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. Review of Research in Education, 8, 158-233.

Coe, R., & Fitz-Gibbon, C. T. (1998). School Effectiveness Research: Criticisms and Recommendations. *Oxford Review of Education, 24*(No. 4), 421-438.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition.* Hillsdale: Lawrence Erlbaum.

Corrin, W., Lindsay, J., Somers, M., Myers, N., Meyers, C., Condon, C., Smith, J., & Garcia, S. (2013). *Evaluation of the Content Literacy Continuum: Report on Program Impacts, Program Fidelity, and Contrast.* (NCEE 2013-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Chubb, J. E., & Moe, T. M. (1990). Politics, Markets, and America's Schools. Washington D.C.: Brookings Institution.

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

Cooper, H. (2010). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach, 4th Edition.* Thousand Oaks, CA: Sage Publications Inc.

Creemers, B. P. M. (1994). *The Effective Classroom.* London: Cassell

Dahlkemper, L. (2003). What Does Scientifically Based Research Mean for Schools? *Southwest*

Educational Development Laboratory, XV(1).

Edmonds, R. R., & et al. (1977). Search for Effective Schools; The Identification and Analysis of City Schools That Are Instructionally Effective for Poor Children.

Edmonds, R. R. (1979a). *A discussion of the literature and issues related to effective schooling*. Cambridge, MA: Center for Urban Studies: Harvard Graduate School of Education.

Edmonds, R. (1979b). *Effective Schools for the Urban Poor*. Educational Leadership, V37, n1, pp 15-17, 20-24.

Edmonds, R. R. (1979c). Some schools work and more can. *Social Policy, 9*, 28-32.

Edmonds, R. R. (1980). *Search for Effective Schools*.

Edmonds, R. R. (1981). Making public schools effective. *Social Policy, 12*(2).

Edmonds, R. R., & Frederiksen, J. R. (1979). *Search for Effective Schools: The Identification and Analysis of City Schools Schools That Are Instructionally Effective for Poor Children.*

Gamse, B., Jacob, R. T., Horst, M., Boulay, B., Unlu, F., Bozzi, L., Caswell, L., & et al. (2008). *Reading First impact study: Final report*. (NCEE 2009-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., Doolittle, F., & Warner, E. (2010). *Middle School mathematics Professional Development Impact Study: Findings After the First Year of Implementation.* (NCEE 2010-4009). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Jencks, C., Smith, M. S., Ackland, H., Bane, M. J., Cohen, D., Grintlis, H., Heynes, B., & Michelson, S. (1972). *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.

Kushman, J., Hanita, M., and Raphael, J. (2011). An experimental study of the project CRISS reading program on grade 9 reading achievement in rural high schools. (NCEE 2010-4007). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hanushek, E. A. (1989). The Impact of Differential Expenditures On School Performance. *Educational Researcher, 18(4).*

Hanushek, E. A. (1994). An Exchange: Part II: Money Might Matter Somewhere: A Response to Hedges, Laine, and Greenwald. *Educational Researcher, 23*(4).

Hattie, J. (2009). *Visible Learning A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. New York, NY: Routledge.

Hattie, J. (2012). *Visible Learning for Teachers: Maximizing Impact of Learning*. New York, NY: Routledge.

Hedges, L. V., Laine, R. D., and Greenwald, R. (1994) An Exchange: Part I: Does Money Matter? A Meta-Analysis of Differential School Inputs on Student Outcomes. *Educational Researcher, 23(3).*

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Money Does Matter Somewhere: A Reply to Hanushek. *Educational Researcher, 23*(4).

Hunter, J. E., & Schmidt, F. L. (2015). *Methods of Meta-Analysis Correcting Error and Bias in Research Findings* (Third ed.). Thousand Oaks, California: SAGE Publications Inc.

Jones, M. S., & Ross, E. F. (1994). School Improvement: A Case Study. An Effective Schools Framework for Partnerships in Systemic Reform.

Klitgaard, R. E., & Hall, G. R. (1974). Are there unusually effective schools? *Journal of Human Resources, 74*, 90-106.

Levine, D. U., Lezotte, L. W., & National Center for Effective Schools Research and Development, M. W. I. (1990). Unusually Effective Schools: A Review and Analysis of Research and Practice.

Lezotte, L.W. (1997). *Learning for All.* Okemos, MI: Effective Schools Products, Ltd.

Lezotte, L., Edmonds, R., & Ratner, G. (1974). *Remedy for School Failure to Equitably Deliver Basic School Skills.* Cambridge: Harvard University.

Lezotte, L., & Passalacqua, J. (1978). Individual School Buildings Do Account for Differences in Measured Pupil Performance.

Lezotte, L. W., & Pepperl, J.-A. C. (1999). *The Effective Schools Process: A Proven Path to Learning for All*. Okemos, MI: Effective Schools Products, Ltd.

Lezotte, L. W., & Snyder, K. M. (2011). *What Effective Schools Do: Re-Envisioning the Correlates.* Bloomington, IN: Solution Tree Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis* (Vol. 49). Thousand Oaks, CA: SAGE Publications, Inc.

Luyten, H. (1994). *School effects: Stability and malleability.* Enschide: University of Twente.

Madaus, G. F., Kellaghan, T., Rakow, E. A., & King, D. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review,* 49(2), 207-230.

Marzano, R. J. (2000). A New Era of School Reform: Going Where the Research Takes Us. Aurora, CO: Mid-continent Research for Education and Learning.

Marzano, R. J. (2003). *What Works in Schools*. Alexandria, VA: ASCD.

Marzano, R., Frontier, T., & Livingston, D. (2011). Effective Supervision: Supporting the Art and Science of Teaching. Alexandria, Virginia: ASCD.

National Center for Education Statistics. (2010). PISA 2009 Results. http://nces.ed.gov/surveys/pisa/pisa2009highlights.asp. Retrieved March 12, 2013.

National Center for Education Statistics. (2012). TIMSS 2011 Results. http://nces.ed.gov/timss/results11.asp. Retrieved November, 10, 2013.

Orland, M., Connolly, B., Fong, T., Sosenko, L. D., Tushnet, N. C., Yin, R. K., Burt, J.M., Warner, E., Office of Planning, Evaluation and Policy Development. Policy and Program Studies Services (2008). Evaluation of the Comprehensive School Reform Program Implementation and Outcomes. Third-Year Report: US Department of Education.

Pink, W. T. (1990). Staff development for urban school improvement: Lessons learned from two case studies. *School Effectiveness and School Improvement, 1*, 41-60.

Prenkert, D. (2015). Effective Schools Research: A Meta-analysis of Progress Monitoring and the Impact on Student Achievement. Ball State University.

Purkey, S. C., & Smith, M. S. (1983). Effective Schools: A Review. *The Elementary School Journal*, *83*(4), 427-452.

Randel, B., Beesley, A. D., Apthorp, H., Clark, T.F., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2011). Classroom assessment for student learning: The impact on elementary school mathematics in the Central Region. (NCEE 2011-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Regional Educational Laboratory Midwest. http://www.relmidwest.org/Ask-A-REL. Retrieved March 28, 2014

Reynolds, D., Stringfield, S., Teddlie, C., & Creemers, B. (2002). The Intellectual and Polic Context. In B. C. David Reynolds, Sam Stringfield, Charles Teddlie, and Gene Schaffer (Ed.), *World Class Schools: International Perspectives on School Effectiveness* (pp. 3-14). New York: RoutledgeFalmer.

Reynolds, D., & Teddlie, C. (2000). The processes of school effectiveness. In C. Teddlie, & D. Reynolds (Eds.), *The international handbook of school effectiveness research (pp. 134-159).* London: Falmer Press.

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., & Townsend, T. (2000). An Introduction to School Effectiveness Research. In C. Teddlie & D. Reynolds (Eds.), *The International Handbook of School Effectiveness Research (pp. 3-25)*. New York: Falmer Press.

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J. & Gallagher, L. (2009). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies.

Rossmiller, R. A., Holcomb, E. L., & Wisconsin Center for Education Research, M. (1993). The Effective Schools Process for Continuous School Improvement.

Rowe, K. J., & Hill, P. W. (1994). Multilevel Modeling in School Effectiveness Research: How Many Levels. Paper presented at the International Congress for School Effectiveness and Improvement, Melbourne, Australia.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research.* London: OFSTED.

Samuelson, A. T. (1972). *Improvements Needed in Federal Agency Coordination And Participation in the Model Cities Program*. Washington D.C.

Saunders, W., Goldenberg, C. , & Gallimore, R. (2009) Increasing achievement by focusing grade level teams on improving classroom learning: A Prospective, Quasi-experimental Study of Title 1 Schools. *American Educational Research Journal*, 46, 4, 1006-1033.

Scheerens, J., & Bosker, R. (1997). *The Foundations of Educational Effectiveness*: Elsevier Science Ltd.

Schmidt, F., & Le, H. (2009). Correcting for the Distorting Effects of Study Artifacts in Meta-Analysis. In L. V. H. Harris Cooper, and Jeffrey C. Valentine (Ed.), *The Handbook of Research Synthesis and Meta-Analysis* (Second ed.). New York: Russell Sage Foundation.

Silberman, C. E. (1970). Crisis in the Classroom: The Remaking of American Education. New York: Random House.

Smink, G. (1991). The Cardiff conference, ICSEI 1991. *Network News International, 1(3),* 2-6.

Snijders, T. A., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Second ed.): SAGE Publications Ltd.

Somers, M. A., Zhu, P., & Wong, E. (2011). *Whether and How to Use State Assessments to Measure Student Achievement in a Multi-State Randomized Experiment: An Empirical Assessment Based on Four Recent Evaluations.* (NCEE 2012-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Stedman, L. C. (1985). A New Look at the Effective Schools Literature. *Urban Education, 20*(No. 3), 295-326.

Stringfield, S., & Teddlie, C. (1988). A Time to Summarize: The Louisiana School Effectiveness Study. Educational Leadership, 46(2), 43.

Stringfield, S., & Teddlie, C. (1989). The first three phases of the Louisiana school effectiveness study. In B. P. M. Creemers, T. Peters, & D. Reynolds (Eds.), *School effectiveness and school improvement: Proceedings of the Second International Congress, Rotterdam* (pp. 281-294). Lisse: Swets & Zeitlinger.

Teddlie, C., & Stringfield, S. (2007). A History of School Effectiveness and Improvement Research in the USA Focusing on the Past Quarter Century. In T. Townsend (Ed.), *International Handbook of School Effectiveness and Improvement* (Vol. 1). The Netherlands: Springer

The No Child Left Behind Act. (2001). Washington D.C.

Townsend, T. (2007). 20 Years of ICSEI: The Impact of School Effectiveness and School Improvement on School Reform. In T. Townsend (Ed.), *International Handbook of School Effectiveness and Improvement* (Vol. 1). The Netherlands: Springer.

United States Department of Education (2009). *Race to the Top Program Executive Summary*. Washington D.C.

Appendix A

Search Terms Used[16]

1. Creating high (and appropriate) expectations for all
    - Teach* expect* AND "effect size"
    - Teach* standards AND "effect size"
    - Instruction* expectations AND "effect size"
    - Instruction* standards AND "effect size"
    - Instruction* goals AND "effect size"

2. Developing staff skills at the school site
    - Professional development AND "effect size"
    - Teach* train* AND "effect size"
    - Teach* professional development AND "effect size"
    - "In-service" AND Teach* or train* AND "effect size"
    - Principal professional development OR Admin* professional development AND "effect size"
    - Principal train* OR Admin* train* AND "effect size"

[16] Note: In addition to the search terms listed, the search term "effect size" was added to each search in order to find studies that impacted student learning.

Appendix B

Coding Guide

Coding Sheet Adapted from page 89, 95, 96, 98, 100, 176 and 177 of Harris Cooper's Research Synthesis and Meta-Analysis (2010). Additionally, the moderator variable questions were adapted from the International Handbook of School Effectiveness and Improvement (2007) on page 143. The coding guide was also created jointly with colleague Debra Prenkert.

Key For Coding Spreadsheet:

Column 1: Study Identification Number beginning with 1 for the first study entered along with author's last name

Column 2 (Study Type)
1a=Frequent, personal monitoring
2=Developing and maintaining a pervasive focus on learning
2a=Focusing on academics
2b=Maximizing school learning time
3=Producing a positive culture
3a=Creating a shared vision
3b=Creating an orderly environment
3c=Emphasizing positive reinforcement
4=Creating high (and appropriate) expectations for all
4a=For students
4b=For staff
5=Monitoring progress at all level
5a=At the school level
5b=At the classroom level
5c=At the student level
6=The process of effective teaching
6a=Maximizing class time
6b=Successful grouping and organization
6c=Exhibiting best teaching practices
6d=Adapting practice to the particulars of classroom
7=Involving parents in productive and appropriate ways
7a=Buffering negative influences
7b=Encouraging productive interactions with parents
8=Developing staff skills at the school site
8a=Site based PD (type of PD was listed next to all 8a codes: One-shot PD, Multiple Session PD, Multiple Session PD with feedback given, or other-specify)
- 8a1= content based PD
- 8a2= pedagogy based PD
- 8a3= both content and pedagogy based PD
- 8a4= leadership PD

8a5= other- please list what "other" is
8a6= not specified in study
9=Emphasizing student responsibilities and rights
9a=Responsibilities
9b=Rights

Column 3: School type will be listed as: a= elementary, b= middle school, c= high school, and 4= other (will specify)

Column 4: Insert grade levels included in the study (K-12)

Column 5: Subject areas included, a= Reading, b= other Language Arts, c=mathematics, d=Science, e=Social Studies, f=other (will specify)

Column 6: Free/Reduced Lunch Percentage (0-100)

Column 7: Type of outcome measured a= Standardized Achievement Test (will specify), b= another test measuring achievement (e.g. teacher developed test, textbook test), c=class grades, d=multiple types (will specify), e=can't tell

Column 8: Effect Size

Column 9: Page number where effect size data was found

Column 10: Number of schools and students included in the sample

Column 11: Date when coding was completed

Column 12: Special education 1, Entire student population 2, English Learners 3, population not specified 0

Column 13: Private schools 1, Public schools 2, school type not specified 0

Column 14: Date of Publication of Data

Column 15: Was new money used to implement (yes or no)

Column 16: Type of Study

All studies are assigned an ID number (column 1). The study associated with each ID number can be found in Appendix D.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Kushman et al	8a3-multiple with feedback	C	9	A	-	Stanford Diagnostic Reading Test, 4th Ed.	*.05*	32	49 schools and 4959 students	8/26/14	2	2	4/2011	No	Cluster randomized trial
1. Kushman et al	8a3-multiple with feedback	C	9	A	-	Stanford Diagnostic Reading Test, 4th Ed.	*-.06*	33	49 schools and 4959 students	8/26/14	2	2	4/2011	No	Cluster randomized trial
1. Kushman et al	8a3-multiple with feedback	C	9	A	-	Stanford Diagnostic Reading Test, 4th Ed.	*-.002*	34	49 schools and 4959 students	8/26/14	2	2	4/2011	No	Cluster randomized trial
2. Randel, et al	8a3 multiple	A	4-5	C	46.59	State Test-CO	*-.06 unadjusted*	67	67 schools and 9596 students	8/27/14	2	2	4/2011	No	Randomized Controlled
2. Randel, et al	8a3 multiple	A	4-5	C	46.59	State Test-CO	*.01 adjusted*	67	67 schools and 9596 students	8/27/14	2	2	4/2011	No	Randomized Controlled
3. Bos, et al	8a3 multiple with feedback	B	7	B	-	State Test-CA	-.01	64	52 schools and 17,837 students	9/1/14	2	2	3/2012	No	Randomized Controlled
3. Bos, et al	8a3 multiple with feedback	B	8	B	-	State Test-CA	.01	64	52 schools and 18,180 students	9/1/14	2	2	3/2012	No	Randomized Controlled
3.Bos, et al	8a3 multiple with feedback	B	7	B	-	State Test-CA	.03	65	52 schools and 7699 students	9/1/14	3	2	3/2012	No	Randomized Controlled
3. Bos, et al	8a3 multiple with feedback	B	8	B	-	State Test-CA	.01	65	52 schools and 8098 students	9/1/14	3	2	3/2012	No	Randomized Controlled
3. Bos, et al	8a3 multiple with	B	6	B	-	State Language Proficiency Test-CA	.05	66	52 schools and 2373 students	9/1/14	3	2	3/2012	No	Randomized Controlled

	feedback														
3. Bos, et al	8a3 multiple with feedback	B	7	B	-	State Test Language Proficiency Test-CA	.03	66	52 schools and 3456 students	9/1/14	3	2	3/2012	No	Randomized Controlled
4. Biancarosa, et al	8a3 multiple with feedback	A	K-2	A	46	DIBELS and Terra Nova	.22	22	17 schools and 3513 students	9/3/14	2	2	9/2010	No	Accelerated multicohort, longitudinal, quasi-experimental
4. Biancarosa, et al	8a3 multiple with feedback	A	K-2	A	46	DIBELS and Terra Nova	.37	22	17 schools and 3541 students	9/3/14	2	2	9/2010	No	Accelerated multicohort, longitudinal, quasi-experimental
4. Biancarosa, et al	8a3 multiple with feedback	A	K-2	A	46	DIBELS and Terra Nova	.43	22	17 schools and 3348 students	9/3/14	2	2	9/2010	No	Accelerated multicohort, longitudinal, quasi-experimental
5. Gamse, et al	8a3 multiple with feedback	A	1	A	NA	Stanford Achievement Test 10th Edition	.10	xv	248 schools and 17,757 students	10/1/14	2	2	11/2009	Yes	Quasi-experimental (Regression Discontinuity design)
5. Gamse, et al	8a3 multiple with feedback	A	2	A	NA	Stanford Achievement Test 10th Edition	.04	xv	248 schools and 17,757 students	10/1/14	2	2	11/2009	Yes	Quasi-experimental (Regression Discontinuity design)
5. Gamse, et al	8a3 multiple with feedback	A	3	A	NA	Stanford Achievement Test 10th Edition	.01	xv	248 schools and 17,757 students	10/1/14	2	2	11/2009	Yes	Quasi-experimental (Regression Discontinuity design)
5. Gamse, et al	8a3 multiple with	A	1	A	NA	Test of Silent Word Reading Fluency	.17	xv	248 schools and 17,757*	10/1/14	2	2	11/2009	Yes	Quasi-experimental (Regression

	feedback								students						Discontinuity design)
6. Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.1928	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.1702	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.2308	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.2214	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.6597	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.6328	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.7447	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental

6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.6450	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	1.0121	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
6. . Saunders et al	8a3 and 8a4 multiple with feedback	A	2-5	A, B, C	86	SAT 9	.8834	1021	15 schools and 13,937 students	10/2/14	2	2	12/2009	No	Quasi-experimental
7. Garet, et al	8a3 multiple with feedback	B	7	C	72.5	NWEA-all	.04	50	77 schools and 4528 students	10/5/14	2	2	4/2010	No	Randomized Controlled
7. Garet, et al	8a3 multiple with feedback	B	7	C	72.5	NWEA-Fractions/ Decimals	.03	50	77 schools and 4528 students	10/5/14	2	2	4/2010	No	Randomized Controlled
7. Garet, et al	8a3 multiple with feedback	B	7	C	72.5	NWEA-Ratio/ Proportion	.03	50	77 schools and 4528 students	10/5/14	2	2	4/2010	No	Randomized Controlled
8. Corrin, et al	8a3 multiple with feedback	C	9	A	56.9	Group Reading Assessment and Diagnostic Examination (comprehesion)	.13	107	28 schools and 7365 students	10/7/14	2	2	12/2012	No	Randomized Controlled
8. Corrin, et al	8a3 multiple with feedback	C	9	A	56.9	Group Reading Assessment and Diagnostic Examination (vocabulary)	.13	110	28 schools and 7365 students	10/7/14	2	2	12/2012	No	Randomized Controlled

8. Corrin, et al	8a3 multiple with feedback	C	9	A	56.9	Group Reading Assessment and Diagnostic Examination (vocabulary)	.09	109	28 schools and 7951 students	10/7/14	2	2	12/2012	No	Randomized Controlled
8. Corrin, et al	8a3 multiple with feedback	C	10	A	56.9	Group Reading Assessment and Diagnostic Examination (vocabulary)	.10	109	28 schools and 8514 students	10/7/14	2	2	12/2012	No	Randomized Controlled
9. Balfanz, et al	8a3 multiple with feedback	B	5	C	NA	Stanford 9	.14	53	6 schools and 1628 students	10/10/14	2	2	2006	No	Quasi-experimental (Matched groups)
9. Balfanz, et al	8a3 multiple with feedback	B	6	C	NA	Stanford 9	.14	53	6 schools and 1628 students	10/10/14	2	2	2006	No	Quasi-experimental (Matched groups)
9. Balfanz, et al	8a3 multiple with feedback	B	7	C	NA	Stanford 9	.25	53	6 schools and 1628 students	10/10/14	2	2	2006	No	Quasi-experimental (Matched groups)
9. Balfanz, et al	8a3 multiple with feedback	B	8	C	NA	Stanford 9	.23	53	6 schools and 1628 students	10/10/14	2	2	2006	No	Quasi-experimental (Matched groups)
9. Balfanz, et al	8a3 multiple with feedback	B	5 to 8	C	NA	Pennsylvania System of School Assessment (state test)-mathematics	.24	53	6 schools and 13,637 students	10/10/14	2	2	2006	No	Quasi-experimental (Matched groups)
10. Roschelle, et al	8a3 multiple	B	7	C	NA	Research Team Developed Test	.63	66	73 schools and 1621 students	10/12/14	2	2	10/2009	No	Randomized Controlled
10. Roschelle, et al	8a3 multiple	B	7	C	NA	Research Team Developed Test	.50	66	25 schools and 1048 students	10/12/14	2	2	10/2009	No	Embedded Quasi-experiment

10. Roschelle, et al	8a3 multiple	B	8	C	NA	Research Team Developed Test	.56	66	42 schools and 825 students	10/12/14	2	2	10/2009	No	Randomized Controlled

Appendix C

Studies Included in Double Coding

Assessing the Value-Added Effects of Literacy Collaborative Professional Development on Student Learning (2010)

Increasing Achievement by Focusing Grade-Level Teams on Improving Classroom Learning: A Prospective, Quasi-Experimental Study of Title I Schools (2009)

The Implementation and Impact of Evidence-Based mathematics Reforms in High-Poverty Middle Schools: A Multi-Site, Multi-Year Study (2006)

Discrepancies between the Two Coders and Resolution

There were no discrepancies between the two coders, but a discussion on how to handle multi-year studies that reported effect sizes for each year of implementation did occur. Since the meta-analysis was geared toward measuring the impact that professional development had on student achievement, both coders agreed that when multiple years were reported, the effect sizes from the final year would be used for the meta-analysis. This would allow for the professional development to have maximum impact on student learning. In addition, this approach allowed the meta-analysis to address any potential relationship between length of professional development delivery and student achievement. All three studies above listed effect sizes for multiple years.

Appendix D

Studies Included in the Meta-Analysis

Studies are listed in the order they were coded and the order they appear in the dissertation.

1) Kushman, J., Hanita, M., and Raphael, J. (2011). An experimental study of the project CRISS reading program on grade 9 reading achievement in rural high schools. (NCEE 2010-4007). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

2) Randel, B., Beesley, A. D., Apthorp, H., Clark, T.F., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2011). Classroom assessment for student learning: The impact on elementary school mathematics in the Central Region. (NCEE 2011-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

3) Bos, J., Sanchez, R., Tseng, F., Rayyes, N., Ortiz, L., and Sinicrope, C. (2012). *Evaluation of Quality Teaching for English Learners (QTEL) Professional Development* (NCEE 2012-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

4) Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literary collaborative professional development on student learning. *The Elementary School Journal, 111(1),* 7-34.

5) Gamse, B., Jacob, R. T., Horst, M., Boulay, B., Unlu, F., Bozzi, L., Caswell, L., & et al. (2008). *Reading First impact study: Final report*. (NCEE 2009-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance.

6) Saunders, W., Goldenberg, C. , & Gallimore, R. (2009) Increasing achievement by focusing

grade level teams on improving classroom learning: A Prospective, Quasi-experimental Study of Title 1 Schools. *American Educational Research Journal*, 46, 4, 1006-1033.

7) Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., Doolittle, F., & Warner, E. (2010). *Middle School mathematics Professional Development Impact Study: Findings After the First Year of Implementation.* (NCEE 2010-4009). Washington, DC: National Center for Education Evaluation and Regional Assistance.

8) Corrin, W., Lindsay, J., Somers, M., Myers, N., Meyers, C., Condon, C., Smith, J., & Garcia, S. (2013). *Evaluation of the Content Literacy Continuum: Report on Program Impacts, Program Fidelity, and Contrast.* (NCEE 2013-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance.

9) Balfanz, R., Mac Iver, D. J., & Byrnes, V. (2006). The implementation and Impact of Evidence-Based mathematics Reforms in High-Poverty Middle Schools: A Multi-Site, Multi-Year Study. *Journal for Research in mathematics Education*, 37(1), 33-64.

10) Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J. & Gallagher, L. (2009). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies.

Appendix E

Study Artifacts and Correction for Attenuation

Study Name	Reliability of Assessment Used	Corrected Effect Size Used
1) An Experimental Study of the Project CRISS Reading Program on Grade 9 Reading Achievement in Rural High Schools (2011)	.80	-.005
2) Classroom Assessment for Student Learning: Impact on Elementary School mathematics in the Central Region (2011)	.94	-.027
3) Evaluation of Quality Teaching for English Learners (QTEL) Professional Development (2012)	.85 .85	-.012 .012
4) Assessing the Value-Added Effects of Literacy Collaborative Professional Development on Student Learning (2010)	.85	.506
5) Reading First Impact Study (2008)	.85 .85 .85	.157 .047 .012
6) Increasing Achievement by Focusing Grade-Level Teams on Improving Classroom Learning: A Prospective, Quasi-Experimental Study of Title I Schools (2009)	.85	1.115
7) Middle School mathematics Professional Development Impact Study (2010)	.85	.039
8) Evaluation of the Content Literacy Continuum: Report on Program Impacts, Program Fidelity, and Contrast (2012)	.90 .84	.113 .119
9) The Implementation and Impact of Evidence-Based mathematics Reforms in High-Poverty Middle Schools: A Multi-Site, Multi-Year Study (2006)	.85	.282
10) Integration of Technology, Curriculum, and Professional Development for Advancing Middle School mathematics: Three Large-Scale Studies (2009)	.85 .85	.665 .658