THE RELATIONSHIP BETWEEN IMPROVED TEACHER EFFECTIVENESS AND STUDENT GROWTH

A Dissertation
Presented to
The Faculty of the Education Department
Carson-Newman University

In Partial Fulfillment
Of the
Requirements for the Degree
Doctor of Education
By
Michelle R. Hope

April, 2019
Dissertation Approval

Student Name: Michelle Hope

Dissertation Title:
THE EFFECT OF IMPROVED TEACHER EFFECTIVENESS ON STUDENT GROWTH

This dissertation has been approved and accepted by the faculty of the Education Department, Carson-Newman University, in partial fulfillment of the requirements for the degree, Doctor of Education.

Dissertation Committee:

Dissertation Chair: Dr. Steve A. Davidson

[Signature]

Methodologist Member: Dr. P. Mark Taylor

[Signature]

Content Member: Dr. Tony L. Dalton

[Signature]

Approved by the Dissertation Committee: April 11, 2019
Abstract

The purpose of this quantitative study was to explore the impact of instructional coaching on student growth, with particular focus on the relationship between improved teacher effectiveness (occurring after participating in a generalized coaching model) and student growth in math and/or reading in K-8 schools in an urban school district in the Southern United States. The study sample for this research was predetermined in that all teacher participants had opted to receive weekly or bi-weekly generalized, district-level coaching in addition to whatever support was already being provided by their school’s instructional leadership team, and so appeared on the list of teachers receiving support that was requested from the district’s research department. This included 77 teachers who received coaching support in 2016, and 79 teachers who received coaching support in 2017. Teacher Effectiveness here was determined by the score (1-5) the teacher received on the district observation rubric (before and after coaching), while student growth was determined by the change in the median of student growth scores from the NWEA MAP test (comparing scores from the Winter testing event-occurring before the first coaching cycle would have started - with the scores from the Spring Testing event). Overall, this study, in an examination of student outcomes for K-8 teachers, who participated in at least one (generalized) coaching cycle, found that changes in teacher observation scores did have a significant impact on student growth scores. Additionally, results from this research study found that the change in teacher observation score (before and after coaching) was significant, and that the degree of change in a teacher’s score was not correlated to the degree of change in student scores. This study also found that generalized coaching does not have a significant impact on student scores assigned to experienced, rather than novice, teachers, even though teacher observation scores were significantly improved for all participating teachers.
Dedication

This project is dedicated to my children, Alyssa and Braedon, who are the reason behind everything I have done and will continue to do.
Acknowledgements

This project was written under the direction and supervision of Dr. Steve Davidson, Dr. Mark P. Taylor, and Dr. Tony Dalton. I thank them for their knowledge, kindness and support through this process.

I would also like to acknowledge the faculty at Carson Newman University, especially Dr. Julia Price, Dr. Patricia Murphree, Dr. Brenda Dean, and the late Dr. Debby Hayes, who were my models for excellence through this program.

I owe additional gratitude to Yolanda Heidelberg, Lori Phillips, and Catherine Diezi for their selfless mentorship, as well as for their love and friendship. I owe the woman and professional I am today in large part to them.

My final acknowledgements are given to my family and friends. To my mother, Anna, to my late father, George, to Kendra Martin, and Rebecca Finlayson, as well as my children, Alyssa and Braedon, and my husband, for every way they have supported, influenced, and loved me in my life and through this journey.
# Table of Contents

Dedication ........................................................................................................ iv

Acknowledgments ......................................................................................... v

Table of Contents .......................................................................................... vi

List of Appendices ...................................................................................... ix

List of Tables, Figures, and Illustrations ...................................................... x

1. Purpose and Organization ....................................................................... 1

   Background of the Study ........................................................................... 1

   Statement of the Problem .......................................................................... 3

   Purpose and Significance of the Study ..................................................... 4

   Theoretical Foundation ............................................................................. 6

   Research Questions .................................................................................. 7

   Limitations ................................................................................................ 7

   Definition of Terms .................................................................................. 9

   Outline of the Document ......................................................................... 11

2. Review of Related Material ..................................................................... 12

   Growth and Achievement in Urban Schools ........................................... 12

   Factors of Student Growth and Achievement ........................................ 15

      Student Level Factors ......................................................................... 15

      School Level Factors .......................................................................... 17

      Teacher Level Factors ........................................................................ 24

   The Significance of the Classroom Teacher ............................................. 28

   Defining Teacher Effectiveness ............................................................... 29

      Teacher Quality ................................................................................... 29

      Teacher Effectiveness ......................................................................... 32

   Improving Teacher Effectiveness ............................................................ 36
Teacher Coaching ........................................................................................................... 39
The Effectiveness of Coaching ....................................................................................... 43
Coaching and Teacher Effectiveness ............................................................................. 44
Coaching and Student Achievement ............................................................................. 46
Additional Issues, Problems and Concerns ................................................................. 47
Study Significance ......................................................................................................... 49
Summary ....................................................................................................................... 50
3. Research Methodology ............................................................................................. 53
   Introduction ................................................................................................................ 53
      Research Questions .................................................................................................. 54
      Population and Sample ........................................................................................... 55
   Instrumentation .......................................................................................................... 56
      MAP Scores ........................................................................................................... 56
      TEM Rubric ............................................................................................................ 61
   Research Procedures ................................................................................................. 65
      Data Collection ....................................................................................................... 65
      Data Analysis ......................................................................................................... 65
   Summary and Conclusion ......................................................................................... 66
4. Data Analysis and Results ......................................................................................... 67
   Introduction ................................................................................................................ 67
   Results ......................................................................................................................... 69
      Question 1: Hypotheses and Analysis .................................................................... 70
      Question 2: Hypotheses and Analysis .................................................................... 73
   Additional Discoveries ............................................................................................ 75
      Coaching and Teacher Effectiveness .................................................................... 75
      Coaching, Teacher Experience, and Student Growth ............................................ 77
   Summary and Conclusion ......................................................................................... 78
5. Conclusions, Implications, and Recommendations .................................................. 81
List of Appendices

Appendix A: IRB Approval ........................................................................................................119
Appendix B: District Approval ..................................................................................................121
Appendix C: Teacher Effectiveness Measurement Rubric ........................................................123
List of Tables and Figures

Tables
Table 4.1: Mean of Student Growth Scores.................................................................71
Table 4.2: Statistical t-test Analysis Comparing Student Scores (2017) ......................72
Table 4.3: Statistical t-test Analysis Comparing Student scores (2016)......................73
Table 4.4: Average Change in Teacher Effectiveness and Student Growth Scores........74
Table 5.1: A t-test Analysis of the Difference between Student Scores.......................84
Table 5.2: A t-test Analysis of the Difference between Student Scores.......................91
Table 5.3: A t-test Analysis of Teacher Effectiveness Scores.......................................92

Figures
Figure 3.1: distribution of teacher ratings in study district....................................... 64
Figure 4.1: Teacher scores before and after coaching for 2017-2018. .........................76
Figure 4.2: Teacher scores before and after coaching for 2016-2017 .........................76
CHAPTER 1
Purpose and Organization

Background of the Study

As schools and districts come under increasing pressure to meet state and federal targets for student achievement, discussions about what factors impact student achievement have come to the forefront of educational policy decision-making. Recent research suggests that these factors can be divided into four key categories: School factors, Social and Individual factors (including family, intelligence, prior schooling, etc.), Social Incentive factors, and Socioeconomic factors (Wisconsin Education Association, 2018). While all of these factors have a role to play in student and school outcomes, Marzano (2003) places particular emphasis on the impact of school-level factors due to the fact that most of the factors that fall into the other categories are beyond the immediate control of schools and teachers. Drawing on 35 years of research, Marzano ranks these school-level factors in order of their impact on student outcomes beginning with a rigorous and viable curriculum, and followed by challenging goals, engaging lessons, and effective feedback with frequent monitoring (teacher quality), parent and community involvement, a safe environment (school culture), and collegiality and professionalism (school leadership and teacher development). Given the importance of teacher quality on student achievement, Marzano (2003) further defines teacher-level factors to include planning, curriculum structure, sequencing and delivery, questioning, instructional strategies, classroom management, relationships, and teacher mindset.

Marzano’s emphasis on teacher-level factors as being particularly influential on student achievement is well supported by empirical research, which suggests that there is significant “additive or cumulative effect of teacher effectiveness on student achievement” (Sanders & Rivers, 1996, p. 3). Indeed, the University of Tennessee’s Value-Added Research and Assessment Center found that students who were placed with high performing teachers for three years in a row (beginning in 3rd grade) scored in the 96th percentile on the state test in
mathematics by the end of grade 5. Comparatively, those students (with comparable academic histories) who were with three low performing teachers in a row scored on average at the 44th percentile (Sanders & Rivers, 1996). This highlights the significant impact of teacher effectiveness on student outcomes and, since the study sample included students from all backgrounds, can be applied to all students of all backgrounds (Wright, Horn & Sanders, 1997). Research also showed that these effects (both for students of strong teachers and for students of weaker teachers) persisted into later years (Wright, Horn & Sanders, 1997). Indeed, studies repeated in both Texas and Tennessee produced similar results each showing that students in the classes of highly effective teachers consistently out-perform students in the classes of low performing teachers (Jordan, Mendro, & Weerasinghe, 1997). This suggests that teacher quality is a significant factor in both student achievement and continued student growth.

Given the understanding that teacher quality significantly impacts student achievement, school efforts to improve student achievement invariably tend to include a focus on improving teacher quality. This typically involves some form of instructional coaching (Galey, 2016), the popularity and staffing rate of which has more than doubled in the last 15 years (Domina, Lewis, Agarwal, & Hanselman, 2015). Instructional Coaching, though it exists in many forms, is essentially the practice of providing job-embedded support, feedback, and mentoring to struggling teachers through the modeling and guided implementation of research-based instructional strategies. The overall goal of these instructional coaching programs is to not only improve teacher practice but to also improve student outcomes. Here, coaches typically serve in one of three roles: cognitive, organizational, or reform (Galey, 2016). Most teacher-level coaching, however, occurs at the cognitive level (Gallucci, Lare, Yoon, & Boatright, 2010; Marsh, McCombs, & Martorell, 2010). Cognitive coaching, more than any other type of coaching, requires that coaches build trusting relationships with their teachers (or mentees) so that they are able to significantly impact the beliefs, behaviors and teaching practices of the
teachers on their caseload (Galey, 2016). However, even though most building-level instructional coaching is cognitive in nature, the act of coaching also has an automatic organizational role due to the fact that coaches spend time providing, and modeling the implementation of, resources to teachers and schools (Galey, 2016). Similarly, as educational issues become increasingly political, especially regarding factors affecting student achievement, coaches must also assume a reform role. This is because instructional coaches must be “agents of reform” (Galey, 2016, p. 61) as they bridge the gap between policy and practice and serve as a spokesperson for district and state mandates or initiatives (Galey, 2016).

Qualitative and anecdotal support for the positive impact of teacher coaching on teacher practice is well known but has ultimately not been well supported with empirical evidence (Coburn & Russell, 2008; Cohen & Hill, 2001; Garet, Porter, Desimone, Birman, & Yoon, 2001; Supovitz & Turner, 2000). This is likely because it can be difficult to separate out factors when it comes to measuring teacher outcomes. More recently one key study involving a review of 44 causal studies by Kraft, Blazar & Hogan (2016), found that instructional coaching does in fact have a large and positive effect on teachers’ instructional practice. Notably, they also found that this effect was more significant than “almost all other school-based interventions” (p. 21) including student incentives, teacher pre-service training, merit-based pay, generalized Professional Development, data-driven instruction, and extended learning time.

**Statement of the Problem**

As evidence that instructional coaching can lead to improved student outcomes grows (Coburn & Russell, 2008; Cohen & Hill, 2001; Garet, Porter, Desimone, Birman, & Yoon, 2001; Supovitz & Turner, 2000), so does its popularity as a form of job-embedded professional development (Demonte, 2013; Neufeld & Roper, 2003; Poglinco & Bach, 2004). However, while there is growing evidence to support the idea that instructional coaching leads to improved teacher outcomes (Annenberg, 2004; Lockwood, McCombs, & Marsh, 2010), there is only
limited research to show that it impacts student outcomes (Bright & Hensley, 2010; Kraft, Blazar, & Hogan, 2016). Indeed, while Kraft, Blazar & Hogan (2016) found that teacher coaching did have a significant impact on teacher practice (0.58 Standard Deviation across 32 studies), they also found that coaching actually only slightly improved student achievement on standardized tests (0.15 standard deviation across 23 studies). This estimation, however, only applies to reading classrooms since 75 percent of the studies examined were in reading classrooms (the effect was 0.11 SD in the two science classrooms, and 0.02 across four math classrooms- neither of which was significant) (Kraft, Blazar, & Hogan, 2016). Similarly, the small sample sizes in these studies, as well as the large variation in type and definition of coaching, makes it difficult to generalize the results, suggesting once again that research connecting teaching coaching with improved student outcomes is limited. This is significant given that the goal of instructional coaching programs is to bring about improvements in student outcomes not just teacher outcomes. In addition, the fact that instructional coaching programs call for significant investments by districts and schools, both in terms of financial resources and human capital, means that issues surrounding the efficacy of instructional coaching programs remain in the forefront.

**Purpose and Significance of the Study**

As our nation leaves behind the *No Child Left Behind Act* (NCLB) (2001) so goes with it the assumption that a “Highly Qualified” teacher is a “Highly Effective” teacher. Indeed, while “Highly Qualified” status, now obsolete with the adoption of ESSA, was defined as “those who hold at least a bachelor’s degree, are fully licensed or certified by the state in the subjects they teach, and can demonstrate competence in the subjects they teach” (Tucker & Stronge, 2005, para.8), there is increasing evidence that these factors alone “are insufficient for teacher effectiveness” (Tucker & Stronge, 2005, para. 8). Rather, teacher effectiveness is defined as being characterized by the careful interplay of a series of complex factors including planning,
organization, data analysis, relationships, and instructional strategy. Specifically, effective teachers are those who are able to “envision instructional goals for their students, then draw upon their knowledge and training to help students achieve success” (Tucker & Stronge, 2005, para. 8).

Given the research on the relationship between teacher effectiveness and student achievement, it is no wonder that schools and districts are seeking ways to improve teacher outcomes, with the overwhelming majority turning to job-embedded professional development programs like instructional coaching as their initiative of choice. The underlying assumption here is that by improving the quality of teacher practice, schools can improve student achievement (Cohen & Hill, 2000; Kennedy, 2016; Scher & O’Reilly, 2009; Weiss & Miller, 2006). Given the limited research to support the relationship between improved teacher outcomes and improved student achievement, this study examined the possible relationship between improved teacher effectiveness (occurring as a result of structured, generalized coaching) and student growth, including whether the degree of teacher improvement is correlated to the degree of student growth, in K-8 classrooms in an urban school district in the Southern United States. For this, teacher observation of practice scores, as determined by the study district’s observation rubric (Teacher Effectiveness Measurement Rubric or TEM Rubric—see Appendix C) were examined for a group of teachers before participation in a coaching model and then again following participation in at least one cycle (typically 9-15 weeks) of coaching. Student growth was measured by using student scores (as assigned to teachers) on the Northwest Education Association (NWEA) Measures of Academic Progress (MAP) test (as the median of all student percentile growth scores assigned to the teacher) before participation in a structured coaching model (looking at growth from Fall to Winter) and then again after at least one complete coaching cycle (looking at student growth from Winter to Spring). This study assumed that students show academic progress at a steady rate across the school year and acknowledges that
some teachers may have needed more than one full cycle of coaching before meeting their individualized coaching goals. To ensure increased reliability and validity of the its findings, this study examines these data for two separate groups of participating teachers, one each from two academic school years: 2016-2017 and 2017-2018.

As schools and districts become more concerned with improving student outcomes by improving teacher outcomes, there is increased emphasis on the importance of instructional coaching in the K-8 setting. This is especially so under ESSA (2015), which places special emphasis on student achievement and growth for all sub-groups, and on teacher professional development (both for instructional strategies and for content specific strategies). Since this study explored whether, and the degree to which, improved teacher outcomes leads to improved student outcome, it has significant implications for school improvement efforts.

**Theoretical Foundation**

The foundational theoretical perspective for this study is rooted in Behaviorism. Behaviorism (specifically that described by Watson, 1913) operates under the assumption that all human behavior (with an emphasis on observed behaviors) is learned. In particular, this theoretical framework assumes that there is a relationship between teacher behaviors and student outcomes and is based on the "process/product paradigm" with the idea that by changing teacher behaviors (the process of teaching) we can change student behaviors/outcomes (the product of learning) (Dunkin & Biddle, 1974). This relates specifically to Applied Behavioral Analysis (ABA), which is defined as, “the science in which tactics derived from the principles of behavior are applied systematically to improve socially significant behavior, and experimentation is used to identify the variables responsible for behavior change” (Cooper, Heron, & Heward, 2007, p. 20). When applied to the process of changing teacher behaviors (the processes), Anderson, Caswell, and Hayes (1994) stressed the importance of “feedback, reinforcement, and controlled conditions for training” (p. 212). Since this comprises the basic conditions for instructional
coaching, one can conclude that coaching itself is grounded in behavioral principles (Kretlow and Bartholomew, 2010). This is especially so given the social nature of coaching and the application of specific and actionable feedback on, and modeling of, teaching behaviors within the environment where they take place (Joyce & Showers, 1980; Killion, 2012; Showers & Joyce, 1996).

In addition to behaviorism, this study also considered the conceptual application of various theories of learning including that by Robert Marzano (2007), which itself is based on the original work by Madeline Hunter (1982). These frameworks for effective teaching are inherently grounded in Behaviorism in that they each prescribe a set of teaching practices identified as leading to improved student outcomes. These elements form the basis of the Teacher Effectiveness Measurement Rubric used by my district and by this study to measure teacher effectiveness.

Research Questions

This study addressed the following research questions:

1. **Research Question 1**: Is there a significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores of teachers after participation in a structured coaching model?

2. **Research Question 2**: Is the degree of teacher growth in observation score (as a result of participation in a coaching model) correlated to the degree of student growth on the MAP test?

Limitations

The following exist as possible limitations to the study:

Use of the TEM rubric to measure effectiveness. In this study, the instrument used to measure teacher effectiveness was a district-level teacher effectiveness measure (TEM) rubric.
Even though observers are required to attend, and pass, norming training exercises, actual implementation and application of the rubric remains interpretative and may be considered subjective. This is especially so if a teacher’s observations are all conducted by the same administrator (which is not meant to happen but cannot be ruled out). In addition is the fact that the TEM rubric is only used by my district (most districts across the state use TEAM), which means that, even though the rubric is based on the components of effective instruction outlined in various theories of learning, there may be some limitation to how the results of the study transfer to other schools and districts.

**NWEA MAP growth scores as a measure of student growth.** In order to get multiple measures of student growth across a variety of grade levels, the NWEA MAP test was used as a source of data. This test, which is administered three times a year in my district, while a good indicator individual student growth and teacher performance, and despite being both nationally normed and aligned to the ACT, is primarily an adaptive test. This means that while it can show the level of student growth, and whether or not students meet their individual goals (determined using normative data), it does not show student proficiency with grade level material. As a result, it would be difficult to generalize the results of this study to apply to student performance or achievement on state tests.

**Identification of Participants.** This study uses data from teachers who participated in a formal coaching model in the study district. While all teachers, in theory, receive some sort of evaluation-based coaching and feedback, this study focuses on those teachers who received additional coaching from district-level coaches. Participation in this coaching model is voluntary and teachers would have had the option to “opt in” or “decline” this additional layer of support. Since not all principals accept this support for their teachers, and not all teachers who are offered support “opt in”, participation in the model, and subsequent inclusion in this study, may indicate that these teachers possess a common mindset, which would serve as a confounding variable. In
addition, since not all teachers will have received coaching from the same support provider, and because there is no way to control for additional factors (such as quality and frequency of school-based coaching, teacher qualifications, etc.) it is possible that any growth in teacher score, or the absence of growth in the score, may be the result of factors other than coaching. However, since the overall aim of the study was to look at what happens to student outcomes when teacher outcomes improve, this limitation should not impact the overall reliability and validity of the study.

**Definition of Terms**

The following definitions for key terms apply to the study:

**Student Growth.** Student Growth in this study refers to how much growth a student shows on a test (in this case the NWEA MAP test) between testing events (NWEA, 2017). It is important to note that the emphasis of this study is student growth, not achievement. This is because there is significant evidence to suggest that growth measures are a much fairer and more accurate way to evaluate schools and teachers (Petrilli & Churchill, 2016), primarily because they control for various issues including background and prior achievement (Petrilli & Churchill, 2016). In addition is the fact that current data shows that an increasingly large percentage of students in urban schools are achieving below state proficiency levels. This means that teachers have an ethical obligation to assess student progress and grow students towards proficiency no matter where they begin.

**Teacher Effectiveness.** There are numerous and varying definitions of Teacher Effectiveness. Generally speaking, Teacher Effectiveness is thought to be comprised of a careful interplay of complex factors including planning, organization, data analysis, relationships, and instructional strategy. Specifically, effective teachers are those who are able to “envision instructional goals for their students, then draw upon their knowledge and training to help students achieve success” (Tucker & Stronge, 2005, para. 8). This is a shift from traditional definition of teacher effectiveness that focused first on teacher qualifications and then on the
typical behaviors associated with teaching. Now, rather than focusing exclusively on the teacher, contemporary definitions of teacher effectiveness trend towards focusing on the act of learning (Varlas, 2009). Similarly, and given the increasingly political role of the classroom teacher in advocating for policy implementation, teacher effectiveness is often associated with how teachers impact classrooms, schools, and their colleagues as well as how they contribute to other important outcomes for students (Goe, Bell, & Little, 2008).

Structured Teacher Coaching. Many definitions of coaching exist especially due to the fact that there are so many different types of coaching. However, at its core, “coaching is characterized by an observation and feedback cycle in an ongoing instructional or clinical situation” (Joyce & Showers, 1981, p.170). Coaches are typically peers (specifically defined as fellow educators) who are experts in the field as evidenced by their own personal use of best teaching and coaching practices (Sailors & Shanklin, 2010). Here, the practice of coaching is essentially an act of professional development whereby teachers are engaged in the process of modeling and practice as a way to see “how and why certain strategies will make a difference for their students” (Russo, 2004, p. 1). Notably, coaching can be both responsive (in helping teachers reflect on their practice) and directive (centered around a cycle of feedback) (Ippolito, 2010), and coaches tend to take on many roles. For the purposes of this study, teacher coaching was viewed as a “key lever” in improving teacher practice and was centered on a type of collaborative coaching. In this type of coaching, coaches and teachers engage in “professional dialogue” (Loftouse, Leat, Towler, Hall, & Cummings, 2010) that focuses on developing specific skills and classroom practices as a way to improve teaching. More specifically, collaborative coaching is considered to be individualized, sustained (occurring for at least one full coaching cycle), relevant (occurring within the context of the teacher’s own classroom), research-based (being facilitated by a considered “expert”), and focused (aimed at improving practice). This study specifically drew data from teachers who were participating in a structured, generalized (i.e. not subject specific) coaching model with district
level coaches. All teachers who participated in the district-level coaching program between 2016 and 2018 were included if they taught reading or math (or both) in a K-8 classroom; teachers who received coaching during the study time period but did not show growth in teacher observation score, or who exclusively taught Social Studies or Science were excluded (since there is no consistent measure of student growth for students in these subjects in the study district), as were those who taught grades 9-12 (since those grades are outside the scope of this study).

Outline of the Document

This study is organized in five chapters. Chapter I includes the background for the study, a statement of the problem and an assertion of the purpose and significance of the study. Also included are the research questions, possible limitations of the study, and a definition of terms. Chapter II provides the review of previous studies including a review of previous findings on the indicators of teacher effectiveness, the efficacy of coaching interventions, and factors affecting student growth and achievement. Chapter III gives an overview of the methodology, quantitative data analysis procedures, the time period of the study, the type of data analyzed in the study, and offers a statement outlining the target population for the study. Chapter IV reviews the results of the study including a discussion of the possible relationship between teacher effectiveness and student growth, as well as whether the degree of teacher growth as a result of instructional coaching is correlated with the degree of student growth on the NWEA MAP test. Finally, Chapter V offers the study conclusions as well as the implications drawn from the study; there is also a recommendation for further research.
CHAPTER 2
Review of Related Material

Growth and Achievement in Urban Schools

In the 35 years since the publication of *A Nation at Risk* (1983) public education in the United States has been subject to numerous reforms. A vast majority of these reforms have been aimed at improving education in urban schools and have included such initiatives as the closure of failing schools, the replacement of teachers and principals, improvements to curriculum, the implementation of teacher evaluation policies, the increase in the number and importance of standardized tests, and the restructuring of district governance to incorporate vouchers, charters, and privatization (Martin, 2016). Even though these initiatives have been met with mixed success, the rationale behind each is that reform in urban education is key to improving the lives of children in poverty (Peck, 2017).

Urban school districts are characterized by five key traits: being located in an urban area, having a high concentration of Economically Disadvantaged students, having a high concentration of English Language Learners, having a large percentage of minority students, and having a large number of schools that are considered “high need” by the State Department of Education (Russo, 2004). Unlike suburban and rural school districts, urban school districts operate in densely populated areas, serve significantly more students, have higher concentrations of poverty, greater racial and ethnic diversity, larger concentrations of immigrant populations and linguistic diversity, and higher rates of student mobility (Kincheloe, 2004, 2010). While socio-demographics are not necessarily the cause of the challenges experienced by urban school systems, attention to these details helps highlight some of the broader social and economic inequities facing these populations, which ultimately impact work of urban schools. This includes the fact that each year, as more and more affluent families move away from inner-cities, urban schools serve larger concentrations of poor students, racial minorities, and English-
Language Learners. This exacerbates the challenges already present in these schools (Boyd, Lankford, Loeb, & Wykoff, 2005).

In 2015, the University of Washington’s Center on Reinventing Public Education commissioned a study to examine urban education in 50 cities across the United States. They found that urban schools have low achievement, limited proficiency gains, and large numbers of schools in the bottom five percent of schools in the state; that students from low-income households and students of color have limited access to high-performing schools and college preparatory experiences (with White students being four times more likely than Black students to enroll in a top-scoring elementary or middle school); that Black students were twice as likely to receive out-of-school suspensions as White students; that less than 15 percent of all high school students took the ACT/SAT in 30 of the 50 cities; and that less than 10 percent of all high school students took advanced math classes each year in 32 of the 50 cities (DeArmond, Denice, Gross, Hernández, & Jochim, 2015). In addition is the fact that, while an achievement gap exists in almost all urban districts, it is larger in large city districts (Blagg, 2016). Indeed, on the 2015 National Assessment of Educational Progress (NAEP) test, the average achievement gap between Black and White student scores was 20 percent larger in large city districts (25 percent larger between Hispanic and White students). This suggests that, despite widespread reform, especially as part of the No Child Left Behind Act (NCLB) (2001), there is still work to be done. Part of the issue is that national reforms like NCLB do not account for the unique challenges of urban education (Ginsberg, 2004). One example of this is seen in the fact that urban school districts often have high drop-out rates, meaning that the cohort of students taking the test changes each year, making it difficult to diagnose and track challenges from year to year. This problem is confounded further by high teacher turn-over rates, which prevent schools from accounting for teacher level factors from year to year, and from accurately comparing scores across schools and within schools. Additional issues include the fact that there are often limited
options for students attending “failing schools” (since so many schools within the community are failing and over-crowded) and that, when available, parents often do not take advantage of opportunities to transfer to other schools. This suggests a broader social issue at work, that of parental education and resource availability. Indeed, the fact is that children attending suburban schools (where property taxes are often almost twice or even three times that of urban public schools and parents are more educated), and children attending private schools (where parents have more disposable income and spending can be as much as ten times that of urban schools) are likely to succeed regardless of reform efforts (Ginsberg, 2004). This is because education is closely connected to social issues and politics (Ginsberg, 2004).

Even though previous educational reforms like NCLB have largely failed to deliver on their goal of improved student achievement in all schools, there is still evidence that the implementation of certain reform efforts leads to student growth and achievement. In particular, in a study of the effect of various turnaround strategies on student performance on the National Assessment of Educational Progress (NAEP) test, Casserly (2012) found that urban school districts can improve math and reading performance through six major reform strategies: strong leadership, robust accountability mechanisms, a coherent curriculum, professional development programming, districtwide support, and quantitative assessment (Cassely, 2012). Evidence for the impact of these reform efforts is seen in districts across Atlanta, Charlotte, and Boston. Each of these districts implemented district and/or building level reform that was well-defined, sustained over time, and supported by district level coaches (Casserly, 2012); each demonstrated improvements to student achievement levels; and each approach to reform shared these six key traits. In addition, a strong and well-communicated vision for success was key in all three districts and was carried into system-wide goals that held all stakeholder members accountable for implementation and success (Casserly, 2012). In keeping with best practice research on school-based reform, all three districts also included coherent, aligned, instructional programs
(especially for literacy) that defined clear expectations for high quality instruction, and supported their instructional programs with frequent professional development and/or instructional coaching. Each of the three districts also designed specific strategies and structures to ensure that their reforms were supported and implemented districtwide through extensive targeted monitoring. In addition, assessment data were used to both measure student learning and to adjust and align resources and support (Casserly, 2012).

Factors Affecting Student Achievement and Growth in Urban Schools

There a number of factors that affect student achievement and growth and, while a number of these factors remain outside of the larger control of the school community (including income level, birth weight, home nutrition, parent education, neighborhood violence, etc.), there are some student level, school level, and teacher level factors that can be controlled. These are discussed as follows.

Student Factors

Student-level factors are those factors individual to students. While there are a number of student-level factors thought to affect student achievement, two key factors- student behavior and limited access to resources— are thought to have the most significant impact on student achievement in urban schools.

Student Behavior. The fact that students in urban areas experience conflict with their community (Ford, 2009) often causes them to be violent and/or defiant, which can inhibit academic progress (Ford, 2009). Student misbehavior in the classroom not only increases the stress levels of students and teachers, but also disrupts the learning environment. In particular, misbehavior changes the dynamics of the classroom, shifting the focus from academics to behavior leading to decreased instructional time, which negatively impacts student outcomes. This emphasizes the idea that student behavior, or more specifically the manner in which teachers are able to manage student behavior, can essentially distinguish an effective classroom
from an ineffective one (Reglin, Akpo-Sanni, & Losike-Sedimo, 2010). In this way, student behavior is a prevailing topic of interest for schools because how students conduct themselves and feel about school has an impact on their overall achievement and performance (Sailor, Zuna, Choi, Thomas, McCart, & Roger, 2006). More specifically, disruptive behavior is considered a growing problem and is identified as an indicator for school performance (Bloom, 2009).

**Limited Access to Resources.** Students who live in urban communities also have limited access to necessary resources (Gold, 2007). This is connected to the idea that a student’s socio-economic status can significantly affect their academic success (*The New Jersey Department of Education*, 2010), and that family background and/or availability of resources at home can explain these achievement differences (Coleman, 1998). One example of how a lack of resources at home affects student outcomes is seen with a recent study on the impact of technology in the home. In a report released by the *Nation’s Report Card* in 2015, researchers found that students (about 20 percent of the total population tested) who did not have access to a computer at home scored lower on the 2015 NAEP mathematics and reading assessments compared to those students who did have access. About half of those students attended urban schools where they did not have access to a computer at school either and, with the exception of the eighth-grade students, students who did not have access to computers at school or home, tended to have lower scores than those who did just did not have access at home. As such, urban schools must be willing to provide resources to bridge these gaps at a faster pace than other schools (Evans, 2001). One such resource might be school-meals, which, in bridging the nutritional gap experienced by children in poverty, can help control for socio-economic status and lead to improved outcomes for all students. Evidence for this was seen in a recent study by Schwartz and Rothbart (2017) who found that when a free school meal was extended to all students regardless of economic status, increases in academic performance by as much as 0.059 standard deviations in math and 0.083 in ELA for non-poor students, and 0.032 and 0.027 standard deviations.
deviations in math and ELA for poor students, were observed.

**School-Level Factors**

School-level factors are those factors controlled by the school and determined by the relationships that exist between stakeholders. Factors that have a large impact on academic achievement in urban schools include Stakeholder Relationships, School Climate, School Leadership, and Academic Press.

**Stakeholder Relationships.** A key school-level factor known to affect academic achievement is the relationship between and amongst stakeholders (Gregory, 2000). There are a number of relationships within this paradigm. The first type of relationship is the relationship that exists between the school and the community. Research shows that if community, school, and parental expectations are in line, the student is more likely to meet the expectations of all educational stakeholders, thereby simultaneously achieving the goals of the school and the community (Coleman, 1988, 1990). This means that schools are stronger when they have close relationships with the communities they serve and when those entities work together to develop a shared vision for success. For urban schools in particular, the effectiveness of strong community relationships was seen in a 2011 study of 35 public elementary schools in an urban center in southeastern Virginia. This study found a positive association between community engagement and student achievement in math and language arts, even when controlling for socio-economic status. Interestingly, study findings also reported that community engagement in schools had an overall positive effect on academic optimism (the collective efficacy of teaching competence and teaching task, faculty trust in students and parents, and academic press or optimism) (Kirby & DiPaola, 2011).

Similarly important is the relationship that is developed and nurtured between schools and parents, often referred to as the home-school partnership. The home-school partnership is loosely defined as the “degree to which parents trust staff, have positive interactions with staff,
and feel welcome at the school” (Wood & Bauman, 2017, p.12), and has been shown to significantly impact student outcomes. Indeed, a study by the University of Chicago’s Consortium on School Research reviewed longitudinal data from 100 elementary schools in Chicago and found that strong parent-school relationships are positively linked to student motivation and attendance. In addition, Byrk Sebring, Allensworth, and Luppescu (2010) found that schools with strong parental support and involvement were ten times more likely to show improvement than schools with limited involvement. In particular they found that when schools used effective family engagement practices, students in those schools were ten times more likely to improve their mathematics performance, and four times more likely to improve their reading performance than students attending schools that did not implement meaningful engagement practices (Bryk et al., 2010). This is likely because parental involvement and parental aspirations have been found to be positively associated with students’ emotional engagement and are more strongly associated for low-socio-economic groups than for high-socio-economic groups (Wang & Sheikh-Khalil, 2014).

Another key stakeholder relationship is the relationship that exists between a principal and his or her teachers. In building relationships with staff members, a principal is able to build trust with and among members, which not only allows school communities to develop shared values but can also significantly increase levels of job satisfaction among teachers (Beauchamp & Parsons, 2012). One key way that principals can foster a community of trust is through shared leadership. In this approach to leadership the principal establishes a clear parameter for leadership as a partnership between all members of the school community (Hughes & Pickeral, 2013). Here, the principal serves as a guide or facilitator for the larger “work” of the school, anchoring all community members in a shared vision or purpose for the school. This shared leadership allows for shared accountability and shared responsibility, and typically embraces multiple perspectives and multiple pathways for problem solving (Hughes & Pickeral, 2013).
This is based on the premise that, in order to foster strong relationships, a principal needs to acknowledge that the teachers and other leaders in the building have an extensive knowledge base (Hughes & Pickeral, 2013). Moreover, once principals are willing to embrace the human resources available to them, school can become a place where “students learn more and feel connected to the adults in the school, the adults in the school collaborate and create a sense of belonging through shared responsibility, and inclusiveness is both a vision and practice” (Hughes & Pickeral, 2013, pg. 2). Notably, in order for shared leadership to positively affect student achievement, the principal must model their vision for a positive school climate in their “words, actions, and beliefs, beginning with the climate they maintain between administration and school staff” (Hughes & Pickeral, 2013, pg. 2). Invariably, this modeled vision for a school’s academic climate is what grounds a school in its pursuit for academic excellence, regardless of its population in that it inspires others (students, parents, teachers) to do the same. This implies that at the heart of all stakeholder relationships is effective collaboration.

Perhaps the most significant collaborative stakeholder relationship in schools is the relationship that exists between and amongst teachers (Wong, 2007). Support for this comes from Rivero (2006) who found that collaboration was a common factor in high performing schools. Indeed, they found that high-performing schools had teachers who were more likely to have a collegial approach to decision-making and more willing to share the knowledge and skills needed to help all students reach high academic standards (cited in Wong, 2007). Ronfeldt, Farmer, McQueen, and Grissom (2015) similarly found that when schools put structures and supports in place for effective, structured collaboration, all types of collaboration significantly and positively predicted school-level math value-added outcomes, while all types of collaboration, except that about assessment, significantly predicted school-level reading value-added outcomes. In particular they found that schools with effective collaboration structures had “higher achievement gains in math and reading” (p. 500), and that teachers who participated (and
saw value) in high quality collaboration saw better achievement gains in math than those teachers who were not part of a collaborative team (or who participated in a team that collaborated ineffectively).

School Climate and Culture. Over the past twenty years, there has been an emphasis on the importance of school climate and its influence on academic achievement (Cohen, McCabe, Michelli, & Pickeral, 2009). Specifically, researchers suggest that if school climate is not a primary focus in the school setting, the academic, social, and operational components are negatively impacted. Here, climate is defined as the temporary attitude of individuals in the building and refers to the overall quality of school life (Cameron & Quinn, 2011) created as a result of the relationships between students and staff (Adeogun & Olisaemeka, 2011). According to the Alliance for the Study of School Climate, school climate can be measured by three key factors: attitude and culture, leadership decisions, and faculty interactions (Clifford, Menon, Gangi, Condon, & Horung, 2012). These factors work together to create an environment that challenges students to meet high expectations, supports students and teachers, and fosters school-wide safety (Gregory, Cornell, & Fan, 2012). In this way, school climate is said to consist of variables that include the social environment as well as student sense of academic rigor, student perception of teacher expectations, teacher-student efforts to improve, parental concern for quality, and the desire and efforts of the principal (Fraser, 2012). Moreover, while the literature suggests that positive climate is beneficial for all individuals involved in schools, a negative climate is seen to have the opposite effect, leading to a diminished perception of safety, at risk behaviors, poor student to student and teacher to student relationships, low morale, high teacher turnover, and increased bullying (Nickerson, Singleton, Schnurr, & Collen, 2014). This is because students need to feel safe and have strong positive relationships with their teachers if they are to exhibit academic growth and achievement.
In addition, recent research continues to suggest that school culture and climate, whether positive or negative, can both indirectly and directly affect student achievement (Kraft, Marinell, & Yee, 2016). Indeed, schools with high teacher turnover, a lack of administrative support, and/or inadequate professional development (all factors associated with culture and climate) tend to have lower documented levels of student achievement (Kraft et al., 2016). In particular, “high rates of teacher turnover impose large financial costs on schools and reduce student achievement by undercutting efforts to build capacity and coordinate instruction among staff” (Kraft et al., 2016, pg. 3). Meanwhile, schools with “supportive professional environments,” presumably as the result of a strong building principal and strong community partnerships, are not only “more likely to retain their teachers . . . [but they] also maximize teachers’ and students’ learning opportunities” (Kraft et al., 2016, pg. 3). In these situations, as student achievement increases as a result of positive climate, and teachers engage in on-going professional learning, relationships are strengthened, trust is deepened, and school climate improves even more. This significantly improves student behavior and teacher retention, two key factors in student achievement (Kraft et al., 2016). This is significant because it suggests that teacher effectiveness, long touted as the single most significant factor in student achievement, cannot exist without consideration of the organizational context in which teachers and students work (Kraft et al., 2016). Indeed, teachers cannot teach effectively, and students cannot learn, in environments where they do not feel safe or where they do not feel valued.

**School Leadership.** There is considerable evidence to support the role of leadership in student achievement. Indeed, Leithwood, Louis, Anderson, & Wahlstrom (2004) suggest that “successful leadership can play a highly significant – and frequently underestimated – role in improving student learning” (p.5). Specifically, they make the following two claims: leadership is second only to classroom instruction among all school-related factors that contribute to what students learn at school with the total (direct and indirect) effects of leadership on student
learning accounting for about a quarter of total school effects; and that leadership effects are “usually largest where and when they are needed most” (p.5). In particular, in today’s educational climate, principals are charged with being instructional leaders first. This idea that principals must move away from being operational managers towards something more closely connected with what happens in the classroom is based on the premise that instructional leadership is one of the most significant factors in determining the effectiveness of a school’s academic program (Leithwood et al., 2010). This is undoubtedly because school leaders are responsible for creating the conditions within which the various school variables can have the most impact on learning (Wallace Foundation, 2011). That is, principals are responsible for creating a school culture and climate that is conducive to academic progress. There are, however, a number of tasks that principals must supervise if they are to effectively lead their school’s instructional program, including: shifting the focus of instruction from teaching to learning; forming collaborative structures and processes for faculty to work together to improve instruction; and ensuring that professional development is ongoing and focused toward school (Lunenburg & Irby, 2006). Thus, principals are charged with refocusing their attention away from managerial tasks towards five key dimensions: a focus on learning, collaboration, using data to improve learning, providing support, and aligning curriculum, instruction, and assessment (Fullan, 2010; Marzano & Waters, 2010). This is based on the idea that culture, defined as the shared attitudes, beliefs, customs, and written and unwritten rules that have been developed over time, does not change as a result of top-down mandates but rather as the result of the “specific displacement of existing norms, structures and processes” (Elmore, 2004, p. 11). A strong academic culture then, one capable of cultivating significant academic progress, comes as a result of replacing these “norms” by modeling the desired replacement values and behaviors (namely a vision for academic success) (Elmore, 2004).

**Vision for Academic Success.** A set of well-communicated, rigorous academic
expectations (sometimes referred to as Academic Press), is another key organizational factor for student achievement. This is defined as the extent to which the school is driven by a quest for academic excellence (Hoy & Tarter, 1997) and is characterized by high academic goals, an orderly and serious learning environment, teacher belief in the ability of all students to achieve, and student respect for high academic achievement (Hoy & Miskel, 2005). Four multi-grade studies in particular have demonstrated that academic press is significantly related to student achievement regardless of socio-economic status (Hoy & Sabo, 1998). An additional study by Alig-Mielcarek and Hoy (2005), using structural equation modeling, took this further to suggest that academic press is a “crucial variable” in explaining academic achievement, more so even than principal leadership (which works indirectly not directly to influence student achievement) (Hoy, Sweetland, & Sweetland, 2001). It is worth noting, however, that the efficacy of academic press lies not just in the setting of high expectations, but with a collective belief among stakeholders that each member of the community is capable of meeting those expectations (Smith & Hoy, 2007). In particular, researchers suggest that there is a significant connection between student achievement and the self-efficacy beliefs of students and teachers, and between student achievement and the collective efficacy beliefs about the school (Goddard, LoGerfo, & Hoy, 2000). Collective efficacy is in itself another school-level factor in that the collective judgement of the faculty about the performance capability of the school as a whole has a school-wide impact on student achievement (Goddard et al., 2000; Hoy et al., 2002). Bandura found that schools where there was a strong collective efficacy were more successful academically than those schools who had doubts about their efficacy (these schools, comparatively, achieved little progress or declined) (Smith & Hoy, 2007). Goddard, Hoy, and Hoy (2000) took this further to suggest that collective efficacy is particularly important in elementary schools being both positively correlated with student achievement and able to control for socio-economic status (Smith & Hoy, 2007).
Teacher Factors

Teacher-level factors are those factors determined by the individual teacher but that can be influenced by the school, student behaviors, stakeholder relationships, etc. Factors particularly important for teachers in urban schools include teacher morale and motivation, self-awareness and reflection, knowledge, and teacher expectations.

**Teacher Morale and Motivation.** Most early research on morale defines morale as a feeling or state of mind that involves a mental and/or emotional attitude (Mendel, 1987). Evans (1997) further defines morale as a state of mind determined by an individual’s anticipation of the satisfaction for needs perceived as important factors affecting the overall work environment. In thinking about the specific implications of teacher morale, Mendel (cited in Lumsden, 1998) suggested that low teacher morale is problematic for schools because it can lead to a decrease in productivity, a loss of concern for the subject or the students, alienation from colleagues, depression, missed work days, general fatigue, and burnout.

Given the impact of low teacher morale, many researchers have attempted to investigate factors contributing to teacher state of mind. Evans (1997) found that a perception of low status, low pay, and a lack of professional autonomy were the three leading factors in determining teacher morale. Further, Dinham (1994) found that teacher morale was most influenced by extrinsic factors such as changes to educational policies and procedures, schools having to deal with social problems, the declining status of teachers in society, poor school leadership, and, increased administrative workloads. Either way, researchers agree that teacher morale is one of key factors in determining school success (Bartell, 1990; Cook, 1979; Whitaker, Whitaker, & Lumpa, 2000), noting that schools with high teacher morale often have high student morale, which has a direct impact on student achievement (Keeler & Andrews, 1963; Whitaker et al., 2000).

Part of the reason that teacher morale is so important to the work of schools is because
the concept of teacher morale is linked to the concept of self-efficacy. Bandura (1977) defined self-efficacy as “the conviction that one can successfully execute the behavior required to produce outcomes” (p. 193), which in this case applies specifically to competencies observed within the classroom. Holzberger, Phillip, and Kunter (2013) took this further to define teacher self-efficacy as “beliefs about their capacity to teach their subject matter even to difficult students” (p. 774). This means that teacher success can actually be predicted by each teacher’s “self-efficacy beliefs about their own abilities to successfully perform specific teaching and learning related tasks within the context of their own classrooms” (Dellinger, Bobbett, Olivier, & Ellett, 2008, p. 751). It’s also worth noting that teacher efficacy is not static but can actually improve with experience and with acquisition of new skills. In this way, teacher development programs that aim to improve teacher practice, also indirectly improve teacher efficacy, which ultimately leads to improved student outcomes (Dunst & Bruder, 2014).

Self-Awareness and Self Reflection. Nowhere is teacher self-awareness and self-reflection more critical than in urban schools. Indeed, teachers in urban schools often experience two competing desires: the desire to meet students’ learning needs in an individual, personal manner, and the desire to implement a system that requires uniform conduct, treatment, and outcomes (Weiner 1993; 1999). In addition, in order to be truly successful, teachers of urban students need to reflect on their own belief systems and assumptions, especially in instances where their social backgrounds and experiences differ from those of the students they teach (Weiner 1993; 1999). Olmedo (1997) described a number of activities found to improve the practice of preservice teachers’ self-awareness, including journaling, composing essays that related readings to practice, and participating in weekly discussions focused on expectations and reflections. These activities helped teachers “analyze and reflect on the impact of their misconceptions of teaching and learning in an urban poor school, and to increase their effectiveness later with their own students” (p.46).
**Knowledge.** Knowledge refers not only to content and/or pedagogical knowledge, but also a deep understanding of the issues particular to urban schools (Guyton, 1994). Here, understanding the individual needs of the students in their classes allows teachers to apply different teaching behaviors to different groups of students. This knowledge might include the effects of poverty on learning, an awareness of the resources available in the school and community, and an acknowledgment of the additional bureaucracy in large urban schools. Murrell (2001), referring to teachers in urban schools as “community teachers,” suggests that teachers need to understand the realities that their students come to school with including hunger, anger, fear, illness, conflict, and transience, all of which impact their ability to learn (and in some cases a teacher’s ability to teach). In a sense this means that community teachers need a deep understanding of the extent and effects of poverty (Guyton, 1994). This allows teachers the opportunity to identify needs within their classrooms and have tools available to meet the needs of students and families. There is also a need for teachers to understand how to navigate some of the demands of working in urban districts where the accountability is high, and teachers have to deal with policy, numerous meetings, interruptions, inadequate materials, lack of time, large classes, and concern with test scores (Haberman 1995).

**High Expectations.** Similar to the presence of a school-wide focus on Academic Press, high expectation at the teacher level is critical to the overall success of students. This is grounded in the idea that successful teachers believe that all students can learn, and that they (the teacher) have the personal and professional capacity to help all students learn (Brophy 1999; Zeichner 2003). Part of this is a deep understanding that lowering expectations for urban students is not only ineffective (Guyton, 1994) but also does a disservice to the students (Diffily & Perkins, 2002). Indeed, when a teacher lowers their expectations for their students because of a belief that there is little that can be done given the students’ backgrounds, student achievement lags, and school reform is not possible (Warren, 2002). The problem with this is that teacher expectation
can be difficult to control for given that the practice of teaching itself requires teachers to evaluate the level of their students and scaffold their instruction to meet students where they are. However, while scaffolds are necessary, teachers can inadvertently cause harm by setting expectations based on a judgement, or by failing to remove or adjust scaffolds as students show progress. One example of this is seen in how a delayed response to a question from non-minority students might be perceived as thoughtful consideration, while the same delayed response from a minority, lower-income student might be considered as a lack of understanding, leading to an unnecessary change in practice or lowering of expectations (Brophy, 1999). This negative, though unintentional, teacher effect is thought to account for 10 percent of the variance in student achievement between student sub-groups (Brophy, 1999).

Evidence for the role of teacher expectation on student achievement dates back to the *Pygmalion in the Classroom* study (Rosenthal & Jacobson, 1968). In this study teachers were given incorrect information about the IQ test results for certain students (teachers were told that these high scoring students were on the brink of rapid intellectual growth). In the study the teachers changed their expectations for, and behaviors toward, the “high performing” students, and those students ultimately demonstrated significantly higher academic achievement than their classmates at the end of the year. More recently, Rubie-Davis, Hattie, & Hamilton (2010) similarly explored the role of teacher perception in determining academic expectations for their students. They found that when teachers had preconceived notions about certain ethnic groups of students (such as coming from low-income families where education is not valued), they had lower expectations for those groups. This has significant implication for urban schools with the idea that teacher expectations can impact student achievement of minority groups, which in turn widens the achievement gap between sub-groups. In particular, the danger comes from the fact that teachers are often “unaware of the differential expectations they have set for their students, not recognizing the minor behavioral cues that can negatively affect their students’ achievement.
levels” (Workman, 2012). This means that school leaders and teacher preparation programs have a responsibility to help teachers recognize their own biases and learn how to tailor their instruction to meet the needs of all students without lowering their expectations (or the expectations of the standards).

**The Classroom Teacher as the Most Significant Predictor of Student Achievement**

Over the past twenty years, research has consistently shown that teaching, or more specifically the qualities of an individual teacher’s effectiveness, is the most significant of all factors affecting student achievement (McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997), and that teacher effects on student learning are both cumulative and long-lasting (Kain, 1998; McCaffrey, Lockwood, & Hamilton, 2003; Mendro, Jordan, Gomez, Anderson, & Bembry, 1998; Rivers, 1999; Sanders & Rivers, 1996). In particular is the idea that teacher expertise accounts for 40 percent of the differences in student performance, which is more than any other factor (Ferguson, 1991), and that students who have an effective teacher for at least three years in a row make significant academic gains (Sack, 1999). This is because the difference in impact of a highly effective teacher (in the top one-sixth) is nine months, which is essentially a whole year of learning (Rowen, Correnti, & Miller, 2002). Moreover, low-income students who have an effective (rather than average or ineffective) teacher for four or five years in a row can actually match or outperform high income students in math (Hanushek, Kain, and Rivkin, 2004).

Research into the relationship between teacher effectiveness and student outcomes has far reaching consequences. This is especially so given the numerous innovations and initiatives that have been proposed and funded in the hopes that they will bring about school improvement. The problem is that, in a review of 40 years of educational research, John Goodlad’s (1990) UCLA study found that not only did none of the studied “innovations” improve student achievement, but the only factor actually shown to impact student achievement (one way or the other) was the
classroom teacher. Similarly, the Annenberg Foundation (1998) found that of all the investments put into education, the investments that yielded the best impact on student achievement were those designed to give teachers sustained opportunities to develop and improve their classroom practice (cited in Wong, 2007). This suggests that in order to improve student learning it is first necessary to change the instructional practices (and therefore, quality) of the teachers responsible for delivering the instruction. This is because good instruction is thought to be 15 to 20 times more powerful than family background, income, race, gender, or other school and student factors (Hershberg, 2005) in determining student achievement. Similar research on reforms in urban districts found that “the only reform effort that clearly resulted in student achievement gains had clear instructional expectations, [was] supported by extensive professional development, [and took place] over a period of several years” (cited in Wong, 2007). This suggests that when districts and schools take the time and money to provide structured, sustained training for their teachers, teacher effectiveness will improve, which in turn leads to improved student outcomes.

Defining Teacher Effectiveness

There are a number of studies that link student achievement to teacher effectiveness (Rivkin, Hanushek, & Kain, 2005; Rockoff, Jacob, Kane, & Staiger, 2008) but these studies vary greatly in their definition and consideration of what constitutes “effectiveness.” Indeed, while some research focused on the importance of teacher qualifications and teacher preparation as benchmarks of teacher quality, other research looks beyond teacher qualifications to those competencies believed to precipitate effective teaching behaviors.

Teacher Quality

For many years, especially under No Child Left Behind (2001), the issue of teacher effectiveness was addressed by requiring that all teachers achieve “Highly Qualified” status. This meant that, at a minimum, all teachers had to possess a bachelor’s degree, obtain full state certification, and show demonstrated competency in each core academic subject he or she
teaches (usually by taking a certification test).

Much of the research (Goe, 2007; Rice, 2003; Wayne & Youngs, 2003; Zeichner & Conklin, 2005) on the impact of teacher qualifications and student performance, found that teachers who have an advanced degree (especially in Math and Science) tend to see higher student achievement. This research, however, is limited to high school and is only significant for Math and Science (Ferguson & Ladd, 1996; Goe, 2007; Goldhaber & Brewer, 2000; Rice, 2003; Wayne & Youngs, 2003). Meanwhile, research into the value of advanced degrees at the elementary level is limited, largely because there was a time when teachers of early childhood education were not required to have a degree at all. As a result, much of the early research in that area focused on the impact of a bachelor’s degree on early childhood teaching. On the whole, however, researchers found that educators with a bachelor’s degree were more responsive to children and tended to offer more activities that promote language development and emergent literacy than teachers without a bachelor’s degree (Ackerman, 2005; Whitebook, 2003). That said, it is consistently apparent that a degree alone is not enough to ensure teacher quality at any level and is why states continue to require most teachers to be licensed to teach (private school high school teachers with a degree in their subject are not always required to obtain certification).

In most states, a teaching certification is obtained by completion of an approved teacher preparation program and a passing score on a certification (PRAXIS) test. There have, however, been a number of concerns about the quality of these programs with many who believe that these programs do little to prepare teachers for the realities of the classroom (especially in urban districts) (Levine, 2006). This is because these programs tend to focus on academic or theoretical approaches to education rather than on the realities of education in practice. This research is supported by the fact that there is limited evidence that teacher certification leads to improved teacher outcomes for all students (Rice, 2003). Indeed, while there are studies that support a
positive relationship between teacher certification in mathematics and student achievement in that subject at the high school level, these studies do not show a similar relationship for other academic subjects in high school or for reading or mathematics in elementary school (Goe, 2007; Goldhaber & Brewer, 2001; Rice, 2003; Rowan, Correnti, & Miller, 2002; Wayne & Youngs, 2003). Additional research has examined the possible relationship between teacher performance on state licensing exams and student test scores finding that teacher test scores account for 25 percent of the variation in student average test scores (Ferguson, 1991).

More recent data on teacher quality include measures to control for student and teacher heterogeneity and tries to control for student and teacher variables. The most influential studies on teacher effects on student achievement come from Rivkin et. al (2005) who, using a value-added gains model for students in grade 3-6, found that teacher quality had a large impact on student achievement. Similarly, Clotfelter, Ladd, and Vigdor (2007) found that teacher experience, education, and licensure test scores have positive effects on student achievement, but also found that these effects, though large (relative to socio-economic characteristics) for math, were smaller in reading. Finally, Koedel and Betts (2007), using a value-added gains model to look at student achievement of elementary students in San Diego, also found that teacher quality is an important predictor of student achievement, but added that measured teacher qualifications (experience, quality of undergraduate college, education level, and college major) have little effect on student achievement. Overall, each of these studies is consistent in showing large effects between teacher quality and student achievement but fail to properly control for teacher heterogeneity. This suggests that, despite efforts over the past ten years to quantify teacher quality by emphasizing the importance of teacher qualifications, teacher credentials and or training are not enough to determine teacher quality or to ensure student success (Rockoff, 2004). This idea that teacher quality cannot be quantified has led to an ongoing discussion of the difference between teacher effectiveness and teacher quality.
Teacher Effectiveness

As conversations about teacher quality have shifted away from a focus on teacher qualifications, researchers have begun to examine exactly what behaviors are displayed by “effective” teachers. Early work by on this was conducted by Skinner (1954) who focused on the importance of reinforcement for effective classroom instruction. In particular, Skinner discussed the importance of dividing processes into “very small steps” (p. 94) and explored the various layers of reinforcement and the impact of teacher behavior on student learning. The idea here is that rather than learning by doing, experiencing, or by trial and error, students learn best when their responses to academic stimuli are immediately and appropriately reinforced. Building on this work, in 2007, Robert Mazano proposed a Framework for Effective Instruction that includes a targeted instructional design. This work was further expanded in 2011 to include four key domains of teacher effectiveness: Domain 1: Classroom Strategies and Behaviors; Domain 2: Planning and Preparing; Domain 3: Reflecting on Teaching; and Domain 4: Collegiality and Professionalism. Of these, Domain 1, which contains 41 discrete elements, most applies to specific classroom instructional behaviors. These behaviors include Learning Goals and Feedback (establishing goals, tracking progress), Rules and Procedures (establishing routines and procedures), Interacting with New Knowledge (identifying critical information, previewing new content, chunking, recording and representing knowledge, reflecting), Practice (similarities and differences, revising, homework, review), Generating and Testing Hypotheses, Student Engagement, Teacher/Student Relationships, and High Expectations (including questioning) (Marzano, 2013). Notably, the Marzano research team asserts that effective instruction requires a careful interplay of behaviors within each domain and equal emphasis across domains.

In an attempt to draw connections between the effective execution of these behaviors and student achievement, a number of research studies have been carried out by the Marzano Research Institute. This included 300 experimental/control studies involving over 14,000
students and 300 teachers, across 38 schools in 14 districts. Overall, the research suggested that effective implementation of the recommended behaviors led to an average effect size on student achievement of .42, with some studies reporting effect sizes of 2.00 and higher. In addition, an average effect size of .42 was associated with a 16-percentile point gain in student achievement (Marzano, 2011). Additional correlational studies conducted using state mathematics and reading test data in urban schools in Oklahoma found that 96% of the 82 correlations (with 41 correlations for mathematics and 41 for reading) were found to be positive, with some as high as .40 or more. This .40 correlation translates to an effect size (standardized mean difference) of .87, which is associated with a 31-percentile point gain in student achievement. These studies also aggregated data across the nine design questions in Domain 1 and found that all correlations were positive for this aggregated data; seven of those correlations ranged from .33 to .40, which translates into effect sizes of .70 and higher. Interestingly, high correlations such as these were also reported for the total number of Domain 1 competencies teachers used, where the number of Domain 1 strategies teachers used in school had a .35 correlation with reaching proficiency and a .26 correlation with mathematics proficiency. This means that the more “effective” behaviors a teacher demonstrates, the more effective he/she is.

Marzano’s research is supported by the Wing Institute (2018) who set out to identify the four most critical competencies associated with teacher effectiveness. Drawing on research that reveals teachers have between a 0.30 and 0.42 effect size on student outcomes (more than any other factor) (Hattie, 2009; Wenglinsky, 2002), the Wing Institute proposed that in order to be effective, teachers must utilize an explicit model of instruction that emphasizes the teacher as an active agent responsible for planning, delivering, and monitoring instruction. In particular, they suggest that effective teachers are responsible for delivering instruction using research-based strategies, for monitoring and adjusting teaching strategies to meet the needs of gifted students and those who are struggling, and for continually assessing students. As such they identify
instructional delivery, classroom management, formative assessment, and personal competencies as the four key competencies characterizing teacher effectiveness (Wing Institute, 2018).

**Instructional Delivery.** Following the design for explicit instruction, The Wing Institute (2018) identifies the following key components of effective instructional delivery: teacher selects the learning area to be taught; teacher sets criteria for success; teacher informs students of criteria ahead of the lesson; teacher demonstrates to the students successful use of the knowledge/skills through modeling; teacher evaluates student acquisition of knowledge; teacher provides remedial opportunities for acquiring the knowledge/skills, if necessary; and the Teacher provides closure at the end of the lesson (Archer & Hughes, 2011; Knight, 2012). Key to this approach to instruction is a well-planned lesson (using an aligned, standards-based curriculum) with clear objectives, careful sequencing of material (connecting to prior learning as well as a “big-idea”), and significant opportunities for students to practice and engage with the objective (through writing and/or discussion). The Wing Institute (2018) also highlights the importance of requiring mastery through implementation of rigorous tasks or text. While each of these practices were found to have an impact on student achievement, four key behaviors were found to have a significant impact on student achievement: the quantity of instruction (0.84 effect size), teacher demonstration/modeling (0.82 effect size), differential reinforcement (0.80 effect size), and cumulative review (0.88 effect size).

**Classroom management.** Classroom management is a persistent concern for classrooms across the country in that it has the potential to significantly hinder a teacher’s ability to teach. This impact is often greater in urban school districts or in classrooms with new or pre-service teachers as the lead teacher (Evertson & Weinstein, 2013). Overall, research suggests that effective behavior management can have a 0.52 effect size impact on student achievement (Hattie, 2012; Marzano, 2003), and that an increase of 20 percent in student achievement can be observed when classroom rules and procedures were applied systematically to a classroom with
management challenges (Hattie, 2012). This highlights the importance of effective classroom management practices for creating an environment conducive to learning, including effective rules and procedures that identify expectations and appropriate behavior for students (broken down into school-wide expectations, and classroom-level expectations) (The Wing Institute, 2018). It is also important that schools and teachers employ proactive classroom management practices, build relationships with students and families, and approach discipline restoratively. This includes communicating and delivering consequences consistently and fairly and allowing students the opportunity to “get it right” (The Wing Institute, 2018). The final component of effective classroom management is effective, engaging instruction. This is based on the premise that students are less likely to misbehave if they are engaged in the work of the lesson (The Wing Institute, 2018).

**Formative assessment.** Effective ongoing assessment (formative assessment, progress monitoring, checking for understanding) is essential for student success (Walberg, 1999). A key component of formative assessment is feedback. Hattie (2009) in particular identified feedback as the single most powerful educational tool available for improving student performance, with an effect size ranging from 0.66 to 0.94. Ongoing assessment (exit tickets, questioning, tests, homework, etc.) is particularly important for effective instruction, having a 0.26 effect size on student achievement; interestingly, this effect size grows to 0.70 when the data is displayed graphically in a meaningful way (The Wing Institute, 2018). Similarly, when teachers use the data to plan instruction, the effect size on student achievement is 0.90 (The Wing Institute, 2018).

**Personal Competencies.** Personal competencies (including establishing high but achievable expectations, encouraging a love for learning, listening to others, being flexible and capable of adjusting to all situations, showing empathy, being culturally sensitive, embedding and encouraging higher order thinking, teaching foundational skills, having a positive regard for
students, and being willing and able to collaborate) are critical to teacher success, especially in urban schools (The Wing Institute, 2018). Indeed, quantitative studies suggest that teachers with strong personal competencies have a greater impact on student behavior and achievement with an effect sizes ranging from 0.72 (on student achievement) to 0.87 (on student behavior) (Cornelius-White, 2007; Marzano, Marzano & Pickering, 2003). The problem is that these competencies are not easily influenced or changed by coaching or teacher preparation. In addition, despite the fact that there is evidence that teacher competencies (including personable competencies) impact student behavior and achievement, there is also research that supports the view that particular teacher characteristics cannot explain teacher effectiveness on their own. Indeed, Rockoff, Jacob, Kane, and Staiger (2008) find that student achievement was most impacted as a result of combined variables on teachers’ cognitive and non-cognitive attributes.

**Improving Teacher Effectiveness**

Given the known relationship between teacher effectiveness and student achievement, it is not surprising that efforts to improve student outcomes also include efforts to improve teacher outcomes. This typically involves job-embedded training or professional development. Teacher professional development in particular aims to enhance teacher knowledge and skills in order to create improved classroom practice; this, in theory, leads to improved student outcomes (Hightower, 2011). In particular, research suggests that well-designed professional development that focuses on deepening subject-matter knowledge, helping teachers understand how students learn, connecting what teachers are being asked to do with what they already know, is effective in developing teacher practice (Cohen & Hill, 2000; Desimone, Porter, Gater, Yoon, & Birman, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Wilson, 2009). Akin to this, Learning Forward (2018) offers schools and districts an outline of the characteristics or standards of professional learning that are known to lead to improved teacher practice. This includes learning
that occurs in learning communities, effective prioritizing and monitoring of resources, learning that integrates theories, research, and models of human learning, strong leadership, emphasis on multiple sources of data, ongoing support for implementation, and a focus on outcomes. Three states in particular (Tennessee, Florida, and Arkansas) have seen a dramatic improvement in the quality of their professional development as a result of adopting Learning Forward’s protocols for professional development (which also include instructional coaching). As a result of this shift, teachers in Arkansas reported increased access to professional development with more than fifty percent of teachers receiving more than 17 hours of content-based instruction (Slabine, 2011). Similarly, over time, districts in Florida found that adopting standards for professional learning helped districts “gain a better understanding of how to design, implement, and maintain quality professional development systems [and to] . . . identify and disseminate effective professional development practices that others could adopt or adapt (Slabine, 2011, p.17). In Tennessee, Learning Forward’s professional learning standards were at the center of reform in Memphis, where the local public-school district used the standards to define learning protocols for both teachers and leaders. Memphis in particular saw improvement in every standard of professional learning, with the greatest improvement in data driven decision making (Slabine, 2011).

This work outlining the importance of high-quality professional learning for teachers is supported by more recent research in which 220 fourth-grade teachers from ninety-four schools in six districts and five states participated and were randomly assigned within schools to either the treatment group that received concentrated, standards-aligned professional development or to the control group that did not. They found that teachers who received the aligned professional development scored 21 points higher on a study-administered math test than the control teachers scored. They also found the professional development appeared to positively impact crucial aspects of the study teachers’ instructional practice, particularly relative to the conceptual aspects
of math and the quality of teachers’ math explanations to students (Northern, 2016).

While there is evidence linking effective professional development to improved teacher practice, evidence connecting effective professional development to student achievement is more limited. This is likely because it can be difficult to isolate the various factors known to contribute to student success. Early research examining the features of high-quality professional development showed that, while professional development significantly increased teacher knowledge and desired classroom practice, it did not necessarily translate into improved student outcomes or even in sustainable changes in teacher practice over time (Garet, Porter, Desimone, Birman, & Yoon, 2008). The problem is that, given that professional development affects student achievement in three steps, determining the impact of professional development can be complex. These steps are that, first, professional development enhances teacher knowledge and skills; and second, that these skills improve classroom teaching; and third, that the improved teaching raises student achievement. As a result, if one link is weak or missing, students will not benefit from the teacher’s professional development. This does not, however, mean that students could not benefit if the practice was properly implemented (Yoon, Duncan Lee, Scarloss, & Shapely, 2007). Indeed, though limited in scope, a study of the research by Yoon et al., (2007) suggests that effective professional development has a moderate effect on student achievement (0.54 effect size in mathematics, science, and reading). It is worth noting, however, that these moderate effect sizes are only observed as a result of content-based development programs, and that research suggests that programs that focus mainly on teacher behaviors have less impact on student learning than programs aimed at improving teacher knowledge of the subject and/or how students learn the subject (Kennedy, 1998; Yoon et al., 2007). This all suggests that while professional development alone is important in working to build capacity of teachers, additional supports (for teachers and administrators) are needed if this is to translate to improved student achievement. One such support includes instructional coaching.
**Teacher Coaching**

Instructional coaches who support school-based professional learning are an important component of an effective job-embedded professional-development protocol (Learning Forward, 2011). This is because coaches can serve a variety of roles in the development of teachers, including as resource providers, data coaches, curriculum specialists, instructional specialists, mentors, classroom supporters, learning facilitators, school leaders, and catalysts for change (Learning Forward, 2011). Defined broadly, coaching is a form of professional development that helps teachers “develop and apply new knowledge, make strong plans for instruction and assessment, obtain feedback, refine their practices, and examine results” (The University of Florida Lastinger Center for Learning, Learning Forward, & Public Impact, 2016, p.5). It is important to note, however, that teacher coaching does not replace other forms of professional development but rather is a form of professional development in itself. Either way, effective coaching is grounded a “relationship between an expert and a novice, or colleagues with similar professional experience, who provide support and feedback to an individual or group for the development of one or more job-specific skills” (Hobson, 2012, p. 60). This emphasis on the relationship between the teacher and coach helps ensure that teachers receive individualized support, and is shown to yield a greater improvement in teacher practice than all other forms of professional development (Hobson, 2012).

Two distinct types of coaching have been shown to be particularly effective in helping to improve teacher practice: peer coaching, where colleagues work together to achieve improvements in their practice; and schoolwide or instructional coaching, where designated coaches, often specializing in literacy or math, visit, advise, plan with, and demonstrate effective practices for teachers across a school, (The University of Florida Lastinger Center for Learning, Learning Forward, & Public Impact, 2016).

**Peer Coaching.** Peer Coaching has been promoted as an essential component of in-
service and preservice teacher training (Joyce & Showers, 2002) and is defined as “a collegial relationship between teachers who provide reciprocal, in-class assistance to each other as they attempt to incorporate new teaching skills, strategies, and approaches toward their teaching” (Neubert & Stover, 1994, p. 7). This means that the majority of peer-coaching occurs between two classroom teachers, where each one has demonstrated expertise in a particular area, which they then share with the other. That said, while most peer coaching is reciprocal (meaning that both parties assume the coaching role at some point within the relationship), it does not always have to be. Indeed, in some instances, peer coaches are full or half release teachers who are tapped to coach new or struggling teachers within their buildings or districts. One such model that uses a non-reciprocal peer-coaching protocol is the Peer Assistance and Review (PAR) program. This program employs exemplary teachers to assist identified struggling veteran and beginning teachers in the areas of subject matter knowledge, instructional techniques (including classroom management) and strategies (Harvard Graduate School of Education, 2018). Unlike other coaching programs that measure success by improved teacher outcomes, the PAR program, which is implemented across the United States (including in the study district), measures success according to successful teacher retention. In other words, the direct goal of the Peer Assistance coaching model is to help teachers retain their employment not necessarily to improve student outcomes (Harvard Graduate School of Education, 2018).

Regardless of reciprocity, a key component of peer coaching is the non-evaluative nature of feedback provided by a colleague who is an expert in the area of focus (Ackland, 1991; Kretlow & Bartholomew, 2010). In particular is the idea that by privileging collegial relationships between teachers, peer coaching promotes collaboration and reduces competition, while developing professionalism within a learning community (Neubert & Stover, 1994). Moreover, by placing emphasis on the teacher as a learner (Fullan, Bennett, & Rolheiser-Bennett, 1990) peer coaching can also work to help teachers implement skills or strategies that
they may have learned in a workshop or class. In this instance, teachers pair with coaches (or other teachers) who help them implement the strategies learned through practice and feedback. This type of coaching is based on the idea that professional development is only effective if teachers first apply their learning to their everyday practice, and then receive follow-up to help with effective implementation (Joyce & Showers 1980). This, however, is unlikely to yield big impact if it is the only kind of coaching teachers receive because it is essentially non-reflective. Indeed, here, changes in practice occur as a result of a focus on implementing required behaviors rather than as a result of a reflection and refinement of personal challenges (Hargreaves & Dawe, 1990). This suggests that, in order for coaching to be effective, activities must have a deliberate instructional focus determined by the teacher’s individual need rather than by a professional development topic. Here, the goal is to refine teacher practice by collaborating to reflect on areas of strength and challenge before working to implement strategies aimed at improving practice. This typically involves evaluating teacher practice against measures of effectiveness (such as that outlined by Marzano, 2011) and selecting an area for improvement. This type of coaching may also include the teacher and coach working together to conduct action research, to solve problems related to instruction or curriculum design and delivery, or to resolve problems with specific (groups of) students (Robbins, 1991). Regardless of the approach, all types of peer-coaching work to refine, expand, and enhance teacher knowledge of best practices related to the profession of teaching (Robbins, 1991).

**Instructional or School-wide Coaching.** Like peer-coaching, instructional coaching centers on a collegial relationship between two individuals. This time, however, it involves a partnership between an “expert level” instructional coach and new and/or struggling teachers, with the goal of improving teaching and learning. To do this, instructional coaches “collaborate with teachers to get a clear picture of current reality, identify goals, pick teaching strategies to meet the goals, monitor progress, and problem solve until the goals are met” (Knight, 2007). As
such, *The Pennsylvania Institution for Coaching* (2018) defines an instructional coach as someone employed to work with teachers and school leaders to bring evidence-based teaching practices into classrooms. At the heart of this is one on one and small group support delivered as part of a strategic instructional coaching cycle, aimed at increasing student engagement, improving student achievement, and building teacher capacity (The Pennsylvania Institution for Coaching, 2018). In addition, even though the coaching process involves two primary individuals: the classroom teacher and the coach, job-embedded coaching requires additional partnerships between instructional coaches and school-leaders. This partnership is essentially reciprocal in that it involves the coach and principal collaborating on ways to best support individual teachers, with the coach supporting work being done by the principal and the principal supporting work being done by the coach. In some instances, where trust and expertise has been established, coaches can also work to build capacity within school leadership teams by serving in an advisory role and supporting principals in identifying and implementing best practices. In this way, instructional coaching is a collaborative and practical approach to school improvement that aims to improve problems in practice. It is, as such, usually on-going, well aligned, and individualized.

It is worth noting that in order for instructional coaching to be effective, especially given the highly political nature of public education, coaches must have credibility. In particular, coaches are called to be experienced, accomplished, and trust-worthy, with a profound ability to manage several priorities at once (The Pennsylvania Institution for Coaching, 2018). This is because coaches are often called on to serve a variety of roles. Key to success in this role, however, is not dependent on the skill-level or experience of the coach, but on his or her ability to establish trusting relationships. It is this relationship that enables coaches to successfully deliver and support professional development sessions, identify classroom-level challenges, recommend strategies for improvement, and see change as a result of those activities. This is
because teachers (and leaders) are more likely to ‘buy-in’ to recommendations if they trust that the coach knows what they are talking about, has the capacity to see the recommendation through to fruition, and will support them through implementation, even when it is difficult.

**The Effectiveness of Coaching on Teacher and Student Outcomes**

Over the last twenty years as districts have explored ways to improve school outcomes, it has become increasingly apparent that high quality professional learning is a “powerful intervention for increasing teaching effectiveness and student learning” (Killion, Harrison, Bryan & Clifton, 2012, p.8). This is especially so for programs that include teacher coaching as part of their professional learning, which has the added benefit of “improving a whole school or department, personalizing professional learning for staff, promoting self-directed professional learning, creating a learning-centered mode of professional dialogue, and building capacity for leadership” (Creasy & Patterson, 2005, p. 20, cited in Killion et al., 2012). The problem is that, while anecdotal evidence for the benefits of coaching programs is abundant, empirical research is both sparse and contradictory. Indeed, while some studies suggest that coaching makes little difference in increasing student achievement or improving teacher practice (Garet et al., 2009), other studies find that coaching has a significant effect on teacher outcomes and (albeit a smaller one) on student outcomes (Killion et al., 2012). These differences in impact, though, likely come as a result of the large variation in how schools and districts view and implement their coaching programs, as well as in the differences inherent in the interactions between coaches and teachers. Either way, there is evidence that well-trained coaches, working within well-structured professional learning program, have impact. Indeed, in their book, *Coaching Matters*, Killion et al. (2012) assert that “coaching can result in increased student achievement, changed teacher practice, and improved school culture if school and district leaders, coaches, and teachers attend to the condition for success” (Killion et al., 2012, p.9). This assumes that when districts ensure that “sound infrastructure, practices, parameters, and relationships” are at work, the likelihood
that coaching will yield its desired outcomes is greatly increased (Killion et al., 2012, p.9).

On the whole, research into the effectiveness of coaching as a form of professional development is divided into its impact on teacher outcomes and its impact on student outcomes. Though limited, this research is consistent in showing that for coaching to have impact, not only do the correct conditions for coaching have to be in place (including well-trained coaches and willing teachers), but that the effects of coaching are generally “greater over time” (Killion et al., 2012, p11).

**Coaching and Teacher Effectiveness**

The premise behind instructional coaching programs is that teachers can essentially be “taught” how to manage the demands of teaching. Specifically, coaching aims to help teachers implement strategies that are known to bring about student success by bringing theory into practice. Indeed, early research into the value of peer coaching on teacher practice found that, while “(d)escription, modeling, practice, and feedback resulted in a 16% to 19% transfer of skill to classroom use. … when coaching was added to the staff development, approximately 95% of the teachers implemented the new skills in their classrooms” (Bush, 1984, p. 197). In 1994, Pierce & Miller similarly found that, when properly trained, peer coaches have a significantly greater impact on the “acquisition of effective teacher behaviors” in novice teachers than traditional supervision (p. 216). This identifies coaching as separate from the typical evaluation and feedback teachers receive from their principal. More recently, Bowman and McCormick (2000) conducted a mixed methods study exploring the differences in a traditional supervision model to a peer coaching model of feedback with focus on the value of peer coaching to preservice teachers in the development of targeted teaching skills, pedagogy, and reasoning. In the study, thirty-two undergraduate elementary education majors in the third or fourth year of an undergraduate program were randomly assigned to the experimental and control groups. At the beginning of the study all participants received the same instruction on the targeted teaching
skills, and the data indicated no statistical significance between the experimental and control groups in their use of these skills during video recorded teaching sessions. The experimental group was then assigned to a peer coach who provided immediate feedback during each of the 12 teaching sessions. Meanwhile, the control group received their feedback (also for all 12 sessions) from the university or the cooperating/mentor teacher. This feedback, however, was based on the candidate’s self-evaluation not the supervisor’s direct observation, and was not immediate (Bowman and McCormick, 2000). Following the experimental cycle, the study reported significant differences in the use of the targeted teaching skills between the two groups, concluding that peer coaching was a more effective way to develop targeting teaching skills in (pre-service and novice) teachers (Bowman & McCormick, 2000).

Rice (2015) also examined the effect of peer coaching on the development of effective teaching behaviors. These behaviors included (a) establishing a student learning objective prior to beginning a lesson, (b) explaining and/or modeling how students can discover, answer, or solve a problem, (c) checking for understanding by asking content-related questions or asking the students to verbally explain/demonstrate the answer/concept, and (d) academic or behavior specific praise statements. However, in this study, which included video-based peer coaching, Binomial logistic regression and ANCOVA analyses indicated no statistically significant differences between the treatment and control groups for the main effects of peer coaching on the development of the targeted effective teaching behaviors.

Studies on the effectiveness of instructional coaching on teacher practice date back to 1982 with the assertion by Showers (1982, 1984) that teachers who receive coaching are more likely to incorporate new teaching practices into their classrooms than teachers who are not coached. This supports the claim that coaching can help improve the efficacy of professional learning in that it increases the likelihood that teachers will actually use their learning. Further research on this was conducted by Joyce and Showers (1995) who found that teachers were 80
percent more likely to use learning from a professional development session in practice when they receive follow-up coaching and support. While these data suggest that 20 percent of teachers may use professional development learning spontaneously and independently, Truesdale (2003) found that after 15 weeks, teachers who did not receive coaching stopped using the new learning, while those who received coaching increased the transfer of the learning to their classrooms (cited in Killion et al., 2012). Kraft, et al. (2016) also examined the effects of instructional coaching on teacher practice using a coded meta-analysis and found that instructional coaching had a significant impact on teacher practice at 0.49 SD. They similarly found that coaching yielded a larger effect size on instruction (0.31 SD greater) when paired with group training sessions.

**Coaching and Student Achievement**

On the whole, studies linking improved student outcomes to teacher coaching focus on subject specific coaching programs. Indeed, a 2006 study by The Learning Network reported steady growth in student achievement over five years in 4th-grade student reading scores when teachers received coaching from literacy coaches, moving from 29% of students achieving proficiency in 1999 to 86% achieving proficiency on state standardized exams in 2004. A similar study of math and science coaching in South Carolina found that teachers who received coaching saw substantial increases in student achievement with a 27% increase in the number of students scoring proficient and advanced (Dempsey, 2007). It is, however, difficult to determine whether these results are due to coaching alone, since most study districts implemented multiple reforms simultaneously. In an attempt to control for other factors, Allen, Pianta, Gregory, Mikami, and Lun (2011) used a randomized control study to examine the effects of a coaching program for secondary teachers on improving student motivation, effort, and achievement. Overall, the study found that students of teachers who received coaching “had a significant net gain relative to the control group … (that) equates to an average increase in student achievement from the 50th to
the 59th percentile” (p. 1035).

More recently, Kraft et al. (2016) offered a theory of change that identified a clear connection between improved instruction and improved student achievement. They found that, across a small sample of 20 studies from 16 research projects that included both outcome measures, the strength of the weighted correlation between averaged effect sizes on instruction and achievement was 0.37 ($p = .16$). In particular, using a weighted linear regression framework, a 1 SD change in instruction was associated with a 0.21 SD change in achievement ($p = .16$). Moreover, ten additional studies allowed the researchers to study the pooled effect on student achievement on state standardized tests, which was measured to be 0.12 SD ($p = .04$, $k = 31$, $n = 10$), although the associated 95% confidence interval includes 0.18 SD. These overall effect size estimates include achievement across reading, math, and science tests in order to provide a broad picture of coaching effectiveness. These studies are not, however, easily generalizable due to the small sample size and because most studies use reading as an outcome measure. This does, however, suggest that instructional coaching aligned to specific content areas (i.e. literacy coaching) has a greater impact on student achievement (0.20SD) than general coaching programs (0.07SD -not significant) (Kraft et al., 2016). Overall, this finding is consistent with a large body of literature documenting the weak relationship between educational inputs (instruction) and outputs (achievement). This, however, does not mean that teacher coaching does not have an impact on student achievement, but rather that the quality of the connected professional development, the quality and frequency of the feedback, and the experience of the teacher, are all compounding factors.

**Additional Issues, Problems, and Concerns**

In considering the relationship between teacher coaching and teacher practice, and between teacher coaching and student outcomes, the research suggests that, while there is evidence that coaching does yield a significant positive effect on teacher practice and on student
achievement, it can be difficult to isolate and control for confounding variables. This includes the frequency of coaching, the type and timeliness of the feedback (Killion et al., 2012), and that it is difficult to control for the type of coaching each teacher receives unless each member of the study receives coaching from the same support provider. Indeed, a triangulation study conducted in the Washoe County (Nevada) School District confirmed a relationship among the frequency of coaching, the program’s duration, and coach experience and student achievement (Taylor, 2008, cited in Killion et al., 2012). This was similarly found in a study by Biancarosa, Bryk, Atteberry, and Hough (2010) who reported that teachers who received high frequency coaching saw a significant increase in their value-added scores, while teachers who received less coaching saw less growth.

Additional variables at work here include the role of the district, the school principal, and the teacher. Taylor (2008) argued that school districts play an “important role in shaping and supporting coaching program effectiveness by: 1) articulating clear expectations related to student achievement and helping schools and coaches focus coaching on achieving those expectations, and 2) providing ongoing support for the selection, professional development, and evaluation of coaching (cited in Killion et al., 2012). In addition, a study of instructional coaching in high schools from 2005-2010 concluded that the frequency of principal and coach meetings was related to both teacher outcomes as a result of coaching, and to student achievement. Notably, the district demonstrating the most significant growth in student achievement had almost daily interactions between the coaches and principals (Sumner, 2011, cited in Killion et al., 2012). Moreover, while teacher engagement in coaching has been shown to impact coaching outcomes, Atteberry and Bryk (2011) suggest that school factors can impact teacher engagement in coaching. Indeed, they found that teachers who felt a sense of responsibility toward the school, fellow teachers, and their students were more open to coaching. Research also suggests that program design can similarly impact teacher engagement. Indeed, a
study in Los Angeles found that direct coach-teacher interactions were more likely to encourage teacher engagement and lead to changes in teacher practice than small group interactions (Rivera, Burley, & Sass, 2004).

Overall, the fact that so many variables contribute to the effectiveness of coaching and to teacher and student outcomes, makes it difficult to fully attribute observed improvements to coaching alone. This is especially so given that schools rarely implement reform efforts in isolation, typically engaging teachers and students in multiple best practice efforts simultaneously.

**Significance of the Current Study**

As schools and districts become more concerned with improving student outcomes there is increased emphasis on the importance of instructional coaching in the K-8 setting. This is especially so under ESSA (2015), which places special emphasis on student achievement and growth for all sub-groups, and on teacher professional development (both for instructional strategies and for content specific strategies) as a means to improve these student outcomes. The underlying assumption here is that, since teacher factors have a greater impact on student achievement than any other factor, by improving the quality of teacher practice, schools can improve student achievement (Cohen & Hill, 2000; Kennedy, 2016; Scher & O’Reilly, 2009; Weiss & Miller, 2006).

Overall, the research points to a definite relationship between teacher coaching and improved teacher effectiveness, and between coaching and improved student achievement, but the significance of this relationship varies greatly across studies. This is largely due to the fact that it is difficult to control for the various confounding variables. As such, the present study looks further at the possible relationship between improved teacher effectiveness (occurring as a result of structured, generalized coaching) and student growth, including whether the degree of teacher improvement is correlated to the degree of student growth, in K-8 classrooms in an
urban school district in the Southern United States. For this, teacher observation of practice scores, as determined by the study district’s observation rubric (Teacher Effectiveness Measure or TEM), were examined before teacher participation in a coaching model and then again following participation in at least one cycle (typically 9-15 weeks) of coaching. Student growth was measured by using student scores (as assigned to teachers) on the Northwest Education Association (NWEA) Measures of Academic Progress (MAP) test (as median percentile growth scores before teacher participation in a structured, generalized coaching model (looking at growth from Fall to Winter) and then again after at least one complete coaching cycle (looking at student growth from Winter to Spring). This study assumed that students show academic progress at a steady rate across the school year and acknowledges that some teachers may have needed more than one full cycle of coaching before meeting their individualized coaching goals. To ensure increased reliability and validity of its findings, this study examined these data for two groups of teachers one from each of the academic school years: 2016-2017 and 2017-2018. Student growth scores were used rather than student achievement so as to help control for student-level factors affecting achievement.

In this way, the current study is able to speak more specifically to the relationship between improved teacher outcomes (teacher growth) and improved student outcomes (student growth), rather than just the generalized relationship between coaching and student achievement. Moreover, since this study explores whether, and the degree to which, improved teacher outcomes leads to improved student outcomes, it has significant implications for school improvement programming efforts.

**Summary**

Research shows that there are a number of school-level, student-level, and teacher-level factors that affect student achievement including: student behavior, student access to resources, stakeholder relationships, school climate and culture, school leadership, academic press, teacher
morale, teacher reflection, teacher knowledge, and teacher expectations. Over the past twenty years, however, research has consistently shown that teacher factors, or more specifically the qualities of an individual teacher’s effectiveness, are the most significant of all factors affecting student achievement (McCaffrey et al., 2003; Rivkin, et al., 2000; Rowan et al. 2002; Wright et al., 1997), and that teacher effects on student learning are both cumulative and long-lasting (Kain, 1998; McCaffrey et al., 2003; Mendro et al., 1998; Rivers, 1999; Sanders & Rivers, 1996).

Given the known relationship between teacher effectiveness and student achievement, efforts to improve student outcomes also tend to include efforts to improve teacher outcomes. This typically involves job-embedded training or professional development. Teacher professional development in particular aims to enhance teacher knowledge and skills in order to create improved classroom practice; this, in theory, leads to improved student outcomes (Hightower, 2011). Research suggests that well-designed professional development that focuses on deepening subject-matter knowledge, helping teachers understand how students learn, and connecting what teachers are being asked to do with what they already know, is most effective at developing teacher practice (Cohen & Hill, 2000; Desimone et al., 2002; Garet et al., 2001; Wilson, 2009). In addition is the idea that teacher coaching, especially when highly individualized, or when paired with high-quality professional development, can lead to significant improvements in teacher practice (Kraft et al., 2016). This change in practice has also been found to lead to improved student outcomes, especially when teachers receive subject or content specific coaching (Allen et al., 2011; Dempsey, 2007). That said, the effect size of the relationship between teacher coaching and student achievement (0.37) is generally smaller than the effect size observed between teacher coaching and teacher practice (0.49) and found to be insignificant (0.07 SD) for the relationship between generalized coaching (rather than 0.20 SD for subject specific coaching) and student achievement (Kraft et al., 2016). It is, however,
difficult to generalize the results from the studies linking teacher coaching with student outcomes due to the fact that the conditions of coaching vary greatly across studies, and because it is difficult to control for confounding variables. What is clear, however, is that well-structured professional development programs, in addition to coaching by well-trained expert coaches, has the potential to significantly impact both teacher practice and student outcomes. This has profound implications for school improvement.
CHAPTER 3
Research Methodology

Introduction

The rationale behind the present study hinges on the underlying assumption that improving the quality of teacher practice leads to improved student achievement (Cohen & Hill, 2000; Kennedy, 2016; Scher & O’Reilly, 2009; Weiss & Miller, 2006). In particular, this study aimed to look further at the possible relationship between improved teacher effectiveness (occurring as a result of structured, generalized coaching) and student growth, including whether the degree of teacher improvement is correlated to the degree of student growth, in K-8 classrooms in an urban school district in the Southern United States. For this, teacher observation of practice scores, as determined by the study district’s observation rubric (Teacher Effectiveness Measure) were examined before teacher participation in a coaching model and then again following participation in at least one cycle (typically 9-15 weeks) of coaching. Student growth was measured by student growth (conditional percentile) scores (as assigned to teachers) on the Northwest Education Association (NWEA) Measures of Academic Progress (MAP) test (median conditional percentile growth scores) before teacher participation in a structured, generalized coaching model (looking at growth from Fall to Winter) and then again after at least one complete coaching cycle (looking at student growth from Winter to Spring). This study assumed that students show academic progress at a steady rate across the school year and acknowledges that some teachers may have needed more than one full cycle of coaching before meeting their individualized coaching goals. To ensure increased reliability and validity of its findings, this study examined these data for two separate groups of teachers from two academic school years: 2016-2017 and 2017-2018. The inclusion of growth scores rather than achievement data here was intentional in that growth calculations allowed the researcher to control for student-level factors affecting achievement. This includes student background, previous educational experiences, family income, etc.
Research Questions

This study addressed the following research questions:

- **Research Question 1:**
  Is there a significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores of teachers after participation in a structured coaching model?

- **Hypothesis 1:**
  There is a statistically significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores after participation in a structured coaching model.

- **Null Hypothesis 1:**
  There is a not a statistically significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores after participation in a structured coaching model.

- **Research Question 2:**
  Is the degree of teacher growth in observation score (as a result of participation in a coaching model) correlated to the degree of student growth on the NWEA MAP test?

- **Hypothesis 2:**
  There is a significant positive correlation between the degree of growth in teacher observation scores and the degree of student growth on the NWEA Map test.

- **Null Hypothesis 2:**
There is no significant positive correlation between the degree of growth in teacher observation scores and the degree of student growth on the NWEA Map test.

**Population and Sample**

The population for this study consisted of all K-8 teachers who were first identified as needing support by the study district, recommended for support by their principal, and who agreed to participate in a collaborative coaching model. This district, identified as one of the largest systems in the nation, serves over 111,500 students with the African American student population accounting for 75.7% of the total enrollment, Hispanic students for 14.2%, Caucasian for 10%, and Asian students for 10%. 82.4% of the student population falls within the economically disadvantaged category (which are students from a household that meets the income eligibility guidelines of less than or equal to 185% of the federal poverty guideline for free and reduced-price meals) (TDOE, 2017). There are 81 elementary schools, 26 middle schools, and 27 high schools; of these: 47 schools are optional (offering specialized or enriched programming) and 18 are I-Zone (Innovation School Zone- a subset of schools that are geared towards turning around underperforming schools). Additional school structures include 13 K-8 schools, eight alternative schools, two “special” schools, one virtual school, four Career & Technical schools, and 45 charter schools. The study district employs 11,500 individuals, 6,800 of which are teachers. Of these, 5,400 are female and 1,400 are male; 175 of these teachers are National Board Certified; 61 percent of teachers are African American; 37 percent of teachers are Caucasian, one percent are Hispanic; and 1 percent are Asian.

To determine the sample size for this study, archived Teacher Effectiveness (TEM) and Student Growth (Measures of Academic Progress or MAP) data for all K-8 teachers identified as not meeting expectations (with a Teacher Effectiveness Score below 2.75/5) from the Fall of 2016 and Fall of 2017 were requested from the district’s research and analysis department. This
totaled 140 teachers for the 2016-2017 school year. From this original list, 63 teachers were eliminated due to being primary teachers of science, social studies, or high school subjects, or due to the absence of observed growth within the coaching cycle. 77 K-8 Reading/Math teachers from 28 individual schools remained. For these teachers, the years of experience ranged from 0.5 (at the end of the study year) to 28 years; 48 of these teachers were in the first five years of teaching. For 2017-2018, data for 110 teachers was received. From this original list, 31 teachers were eliminated due to being primary teachers of science, social studies, or high school subjects (all teachers demonstrated growth following a coaching cycle). 79 K-8 Reading/Math teachers from 23 individual schools remained. For these teachers, years of experience ranged from one year (at the end of the study year) to 24 years; 49 of these teachers were in their first five years of teaching. No teachers appeared in data for both years, but nine schools were served in both the 2016-2017 and 2017-2018 school years.

Description of Instruments

There were two key instruments in this study: the NWEA MAP test and the district’s Teacher Effectiveness Measure Rubric.

MAP growth Scores

MAP Growth scores will be collected from the NWEA MAP test administered to students in the study district three times a year. Scores are presented as the median (for the teacher) of the conditional growth percentile scores assigned to each student. This score represents the level of growth relative to each student’s individual growth goal (or RIT score) measured against a set of national norms. The median is used here rather than the mean given the small size of each group of students assigned to each teacher.

Design and Quality. The Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) system was designed with a specific educational purpose: to provide teachers with valid assessment data. This validity is hinged on six key principles: that the test
should be challenging for a student across all its test items, be economical in its use of student
time, provide a reflection of a student’s achievement that is as accurate and reliable as needed for
the decisions to be made based on its results, consist of content the student should have had an
opportunity to learn, provide information about a student’s change in achievement level from one
test occasion to another as well as the student’s current achievement level, and provide results to
educators and other stakeholders as quickly as possible while maintaining a high level of
integrity in the reported results (NWEA Technical Manual, 2011). It is important to note,
however, that while these design principles are validated at the school level, they do not wholly
encompass true reliability and validity.

One component of NWEA MAP Reliability and Validity applies to the quality of the
individual test items, all of which are evaluated as part of an Internal Quality Review (IQR) to
verify content validity, instructional relevance, and currency. The content validity evaluation
involves ensuring that the item assesses what it is intended to assess and that all parts of the item
are correct, including that any context used in the item is appropriate, the item has one and only
one correct response, and the distractors are plausible and based on student misconceptions; the
instructional relevancy evaluation involves ensuring that the concept assessed is presented in a
way that is consistent with current classroom practice; and the currency evaluation involves
ensuring that the terminology or information in the item is not dated or likely to become dated
(NWEA Technical Manual, 2011). Once this internal quality review has been completed (and all
items verified for plagiarism), the next step in ensuring item quality is with the Editorial Review
(NWEA Technical Manual, 2011). This review involves a more in-depth examination of the
directions, asset, stem, and distractors in order to ensure that the item is aligned to the
specification and is a valid measure of the target concept (NWEA Technical Manual, 2011). In
addition, all items are reviewed for the appropriate reading level, bias and sensitivity, and
Measurement Scales and Calibration. The measurement quality of any test, including the reliability, validity, and precision of its scores, depends largely on the quality of the scale used to measure the variable (NWEA Technical Manual, 2011). In particular, research suggests that on order to be useful, a scale’s unit of difficulty must be a constant size (NWEA Technical Manual, 2011). The NWEA MAP test uses the one-parameter logistic IRT model (1PL), also known as the Rasch model. This model estimates the probability, \( P_{ij} \), that a student, \( j \), with an achievement level of \( \theta_j \), will correctly answer a test question, \( i \), of difficulty \( \delta_i \), and can be expressed as:

\[
P_{ij} = \frac{e^{(\theta_j - \delta_i)}}{1 + e^{(\theta_j - \delta_i)}},
\]

A benefit of the use of an IRT model is that values of achievement levels, \( \theta_j \), and the values of item difficulties, \( \delta_i \), reside on the same equal interval scale (NWEA Technical Manual, 2011). Since this study also used “RIT” (Rasch unIT) scores, it is worth noting that that the MAP test also places value of these achievement levels and item difficulties on the logit metric (an arbitrary scale that is commonly used for academic studies of the Rasch model) (NWEA Technical Manual, 2011). To allow the measurement scale to be easily used in educational settings, a linear transformation of the logit scale is performed to place it onto the RIT (Rasch unIT) scale. This transformation is:

\[
RIT = (\theta_j * 10) + 200
\]

This scale has positive scores for all practical measurement applications and is not easily mistaken for other common educational measurement scales (NWEA Technical Manual, 2011).
Reliability Consistency. The fact that MAP is an adaptive test means that reliability is assessed using methods different from traditional reliability assessments (NWEA Technical Manual, 2011). Indeed, the usual “Test-retest” reliability assessment is not possible because “dynamic item selection is an integral part of the test” (NWEA Technical Manual, 2011, p.55). Similarly, since the difficulty of the item presented depends on the student response to prior items, the test-retest assessment, and even parallel forms reliability, is not always possible, at least not independently. Thus, test-retest reliability of MAP tests is more accurately described as a “mix between test-retest reliability and a type of parallel forms reliability, both of which are spread across several months – a much longer time frame than the typical two or three weeks” (NWEA Technical Manual, 2011, p.55). Specifically, the second test, rather than being the same test, is one comparable to the first, differing only (and maybe only) in terms of the adaptive difficulty level. In this way, both “temporally-related and parallel forms of reliability are framed here as the consistency of covalent measures taken across time.” This is referred to as “stratified, randomly-parallel form reliability” (Green, Bock, Humphreys, Linn, & Reckase, 1984, p. 353, cited in NWEA Technical Manual, 2011) and is best framed in the context of the correlation “between two tests administered from two different but related item pools and those administered twice but from different item pools” (NWEA Technical Manual, 2011). For this study, additional reliability comes from the correlation of MAP scores from tests taken in one term (i.e. Winter) with the (more or less) same students testing the following term (i.e. Spring) since the items for pairs of tests (i.e. Winter and Spring), were drawn from the same item pools (NWEA Technical Manual, 2011). This reliability was repeated for this study.

Similarly, while traditional methods of determining internal consistency depend on all test takers taking the same test, this is not possible with an adaptive test like MAP. Instead, NWEA relies on the marginal reliability coefficient (Samejima, 1994). This coefficient incorporates measurement error as a function of the test score by combining measurement error
estimated at different points on the achievement scale into a single index (NWEA Technical Manual, 2011). This method of calculating internal consistency, $\rho_\theta$, yields results that are nearly identical to coefficient alpha when both methods are applied to the same fixed-form tests and is given by:

$$\rho_\theta = \frac{\sigma^2_{\theta} - M_{s_\theta}^2}{\sigma^2_{\theta}}$$  \hspace{1cm} (3)$$

where: $\sigma^2_{\theta}$ is the observed error variance of the achievement estimates, $\theta$, (the RIT score) and $M_{s_\theta}^2$ is the observed mean of the score’s conditional error variances at each value of $\theta$.

These estimates are made from the pooled results of covalent assessments administered from the same item pools (NWEA Technical Manual, 2011).

**Validity Evidence.** Generally speaking there are a number of factors that can be used to determine validity, ranging from “the adequacy and coverage of a test’s content, to its ability to yield scores that are predictive of a status in some area, to its ability to draw accurate inferences about a test taker’s status with respect to a construct, to its ability to allow generalizations from test performance within a domain to like performance in the same domain” (NWEA Technical Manual, 2011). For the MAP test, most of this evidence comes from the relationships of MAP test scores to state content-aligned accountability test scores (NWEA Technical Manual, 2011), and is generally divided into content, predictive, and criterion validity. Generally speaking, MAP assessment content validity is developed by ensuring that content or state standards are integrated into the test using highly sophisticated computer programming systems, while concurrent validity is expressed in the form of a Pearson correlation coefficient between the total domain area RIT score and the total scale score of another established test designed to assess the same domain area. This essentially answers the question: “How well do the scores from this test that reference this (RIT) scale in this subject area (i.e. Reading) correspond to the scores
obtained from an established test that references some other scale in the same subject area?” In this case, strong concurrent validity is indicated when the correlations are in the mid-.80’s (NWEA Technical Manual, 2011). An additional source of validity evidence for NWEA tests is in their relationship to performance on other tests measuring achievement in the same domain at some later point in time. This form of predictive validity can also be expressed in the form of a Pearson correlation coefficient between the total domain area RIT score and the total scale score of another established test; it answers the question: “How well do the scores from this test that reference this (RIT) scale in this subject area (i.e. Reading) predict the scores obtained from an established test that references some other scale in the same subject area at a later point in time?” In this case, strong predictive validity is indicated when the correlations are in the low .80’s (NWEA Technical Manual, 2011). Finally, the extent to which test scores relate to an external performance criterion provide evidence of an additional form of validity. Criterion-related validity is typically expressed as “the correlation between a test score and a dichotomous criterion such as graduate-not graduate, pass-fail, meet standard/not meet standard, or acceptable-not acceptable job performance” (NWEA Technical Manual, 2011, p.190). For MAP, student performance at the proficient level or above on a state assessment is used as the external criterion, and the relationship between a MAP score and a proficient-not proficient designation on a state assessment is expressed as a point-measure correlation (NWEA Technical Manual, 2011).

**Teacher Effectiveness Measure**

Teacher Effectiveness in this study is measured by the study district’s Teacher Effectiveness Measurement (TEM) Rubric. The rubric measures teacher effectiveness on seven indicators (Objective Driven Lessons; Explaining Content; Appropriately Challenging Work; Content Engagement; Higher Order Thinking Skills; Checking for Understanding; Instructional Time) on a 5-point scale. Teachers are observed multiple times a year with the number of
observations depending on the teacher’s identified “track”. These tracks are: Novice – Educators who are new or educators with no prior year TEM Level of Effectiveness score (receiving four observations); Focus- Educators who are assigned an Instructional Coaching Conversations (ICC) having received either an overall average of Level 1 or 2 the previous school year in observation of practice, portfolio, student survey (TRIPOD), and/or TEM Level of Effectiveness score (receiving three observations); and Proficient – Educators who do not meet any of the previously mentioned categories, neither Novice nor Focus (receiving 2 observations). For the purposes of this study the composite observation score before coaching (pulled from the Fall of the study years) and after coaching (pulled from the Spring of the study years) was used.

Even though the TEM rubric is a rubric specific to the study district, its components are derived from researched Teacher Quality Indicators and was developed as a result of a partnership with the MET Project and Gates Foundation. As a result, many districts use similar indicators to measure teacher effectiveness. The design, reliability, and validity of the rubric is discussed further below.

**Design and Quality.** The direct observation of how teachers teach has been used as a measure of teacher effectiveness for many years, typically with the use of a structured rubric (Danielson, 2013). In particular, The MET Project, supported by the Gates Foundation (2016) showed that when combined with a measure of student achievement growth and student feedback, direct observations of how teachers teach, can be used to create a reliable and valid composite measure of teacher performance (Kane and Staiger, 2012). The foundation used these findings to inform the Intensive Partnership for Effective Teaching protocol, which helped guide a number of districts, including the study district, in working to improve student outcomes by “increasing students’ access to effective teaching” (Stecher, 2016, p.xxv). At the center of the partnership was the development of a “robust measure of teaching effectiveness” (Stecher, 2016, p.xxv). The basis of the study district’s observation rubric was the Washington, D.C., IMPACT
rubric (District of Columbia Public Schools, undated), and uses the same seven indicators of teacher effectiveness (the IMPACT rubric has additional indicators to assess classroom culture and climate that are no longer used in the current version of the study district’s rubric). In both instances, the teacher is observed and rated on all seven indicators (each of which contains 4-8 descriptors) several times during the school year, the first observation of which is announced. The study district uses five (compared to IMPACT’S four) ratings (where the fifth rating includes all descriptors for a level (3) score and all descriptors for a level (4) score). These are: Significantly Above Expectations (5), Above Expectations (4), Meeting Expectations (3), Below Expectations (2), and Significantly Below Expectations (1). These scores are then averaged to give a composite score (Gitomer, Crouse, and Joyce, 2011).

The problem, generally, with an observation rubric is that its effectiveness is wholly dependent on the way it is implemented (Jackson & Mackler, 2016). TNTP (2009) in particular found that in districts using rubrics with a four-point scale, 94 percent of the teachers were awarded one of the top two ratings, and less than 1 percent were rated at the lowest level (Weisberg, Sexton, Mulhern & Keeling, 2009). This failure of observers to capture variation in teacher effectiveness “undermines the potential for observation rubrics as both an evaluation tool and a diagnostic tool to improve teacher practice” (Jackson & Mackler, 2016, p.11). This is why measures of validity and reliability are so critical. It’s worth noting, however, that reliability in this instance does not mean that the teacher’s different observations should be comparable in score since many other factors can impact an individual lesson score (including the students, teacher coaching, time of day, etc.) but rather that two certified observers should score the same lesson in a comparable way, or that one certified individual would score two similar lessons comparably. Similarly, validity here refers to the way in which the rubric is used and the degree of integrity with which it is used.

**Rubric Validity and Reliability.** One key way that both the IMPACT rubric and TEM
rubric get their validity is that there is a substantial observed variability in how raters assign scores and how scores and scoring vary over time and lessons (Casabianca, McCaffrey & Lockwood, 2014). This distribution of scores in the study district is shown in Figure 1 below:

![Figure 1. Distribution of teacher ratings in study district between 2012 and 2016.](image)

The distribution of scores indicates the degree to which the rubric, and its observers, are able to accurately differentiate among teachers. This indicates a positive validity to the utility of the teacher effectiveness measure (Stecher, 2016).

Another measure of rubric reliability and validity is found in the way that observers are trained on the rubric. Ongoing training is necessary to ensure that raters continue to apply the rubric accurately and to provide new raters with the necessary experience (Gitomer, Crouse, & Joyce, 2011). In the study district, all observers are required to attend a certification training session and pass a calibration exercise (in 2018, returning observers were allowed to recalibrate online). Those individuals who do not pass the initial calibration exercise are invited to a second “deep-dive” training before reattempting to calibrate. Passing this exercise is required before permissions to complete (and view) observations are granted, and helps ensure fidelity of the rubric’s use (Stecher, 2016), which implies its validity. Additional reliability of the TEM rubric comes from the fact that at least two different evaluators are required to complete the teacher observations and that no one evaluator can be responsible for more than fifty percent of a teacher’s observations. This helps decrease bias and is important because research suggests that
adding raters improves reliability more than increasing the number of observations by the same rater (Hill, Charalambous & Kraft, 2012).

**Research Procedures and Time Period of the Study**

**Research Design and Collection**

The research design for this study uses a quantitative causal comparative research design, which requires the researcher to select two comparison groups. I also considered the bivariate correlational design, however, since my research is to identify the differences (if any) between the variables, and because correlation is not a predictor of causation (and it implies that the two variables are correlated), this design did not support the purpose of this study.

The study used data that had already been collected by submitting a request to the district’s research department. This “data pull” included a list of all K-8 teachers identified as participating in structured, generalized coaching (appearing on a list of teachers receiving support from the district’s Peer Assistance and Review or PAR program) for the 2016-2017 school year and the 2017-2018 school year. For each of these teachers, TEM observation of practice data (scores of 1-5) from the Fall and Spring of each study year was collected, as well as how their students performed on the NWEA MAP test for the Fall-Winter and on the Winter to Spring assessment for the same academic year. The MAP data appeared as the median of the conditional percentile growth score (1-100) assigned to each student. As an additional point of interest, additional information was also requested (where available) including: teacher years in profession, grade level taught, and school zip code. To ensure confidentiality, once these data were tabulated, teacher names were removed.

**Data Analysis**

This study first looked to examine the difference between student growth scores of teachers before the teacher participated in a coaching model, and student growth scores of the same teacher after they participated in a coaching model. Given that the dependent *t* test (also
called the paired $t$ test or paired-samples $t$ test) compares the means of two related groups to determine whether there is a statistically significant difference between these means, this was determined as the most appropriate statistical test for the study. This is especially so given that a $t$ test requires one dependent variable (in this case, student growth scores) and one categorical variable with two related groups. Indeed, ultimately, a dependent $t$ test is an example of a "within-subjects" or "repeated-measures" statistical test in that the same participants are tested more than once on the same dependent variable.

The study’s second question examined the possible relationship between the degree of teacher growth and the degree of student growth. Since a Pearson’s correlation test is used when you want to find a linear relationship between two independent variables, this was determined as the most appropriate test for this analysis. This is especially so given that Pearson’s is best used for testing in “within groups” studies, which is true of the present study, and will give information both about the magnitude of the correlation and the direction of the relationship.

**Summary**

The purpose of this study was to examine the possible relationship between improved teacher effectiveness, as a result of coaching, and student growth, as well as the correlation between degree of teacher growth and the degree of student growth. A total of 156 K-8 teachers (79 for 2017-2018 and 77 for 2016-2017) from an urban school district in the Southern United States served as the study sample. Data collected included teacher observation data (from TEM observations) and student growth data (from NWEA MAP testing).
CHAPTER 4
Data Analysis and Results

Introduction

Research continues to suggest that teacher effectiveness is a key factor in student achievement and growth (Wright, Horn & Sanders, 1997), outweighing almost all other school level and student level factors thought to influence student achievement. In particular, key research from studies in both Texas and Tennessee demonstrated that students in the classes of highly effective teachers tend to out-perform students in the classes of low performing teachers (Jordan, Mendro, & Weerasinghe, 1997), and suggest that the effects of teacher effectiveness (whether paired with a high performing teacher or low performing teacher) on student achievement persists into later grades.

For the purposes of this study, teacher effectiveness is defined as being characterized by the careful interplay of a series of complex factors including planning, organization, data analysis, relationships, and instructional strategy. Specifically, effective teachers are those who are able to “envision instructional goals for their students, then draw upon their knowledge and training to help students achieve success” (Tucker & Stronge, 2005, para. 8). As a result, given the increasing body of research on the relationship between teacher effectiveness and student achievement, more and more schools and districts are turning to instructional coaching as their initiative of choice. The underlying assumption here is that by improving the quality of teacher practice, schools can improve student achievement (Cohen & Hill, 2000; Kennedy, 2016; Scher & O’Reilly, 2009; Weiss & Miller, 2006). Thus, as evidence that instructional coaching can lead to improved student outcomes grows (Coburn & Russell, 2008; Cohen & Hill, 2001; Garet, et al., 2001; Supovitz & Turner, 2000), so does its popularity as a form of job-embedded professional development (Demonte, 2013; Neufeld & Roper, 2003; Poglinco & Bach, 2004). However, while there is growing evidence to support the idea that instructional coaching leads to improved teacher outcomes (Annenberg, 2004; Lockwood, McCombs, & Marsh, 2010), there is only
limited research to show that it impacts student outcomes (Bright & Hensley, 2010; Kraft, Blazar, & Hogan, 2016). Indeed, Kraft, Blazar & Hogan (2016), in a review of recent research on this subject found that coaching actually only slightly improved student achievement on standardized tests (0.15 standard deviation across 23 studies), and that generalized instructional coaching has a smaller effect on student achievement, at 0.10 SD, than content specific programs at 0.16SD. This is significant given that the goal of instructional coaching programs is to bring about improvements in student outcomes not just teacher outcomes. This, however, does not mean that teacher coaching does not have an impact on student achievement, but rather that the quality of the connected professional development, the quality and frequency of the feedback, and the experience of the teacher, are all compounding factors.

This study examined the possible relationship between improved teacher effectiveness (occurring as a result of structured, generalized, coaching) and student growth, including whether the degree of teacher improvement is correlated to the degree of student growth, in K-8 classrooms in an urban school district in the Southern United States. For this, teacher observation of practice scores, as determined by the study district’s observation rubric (Teacher Effectiveness Measure-see Appendix C) were examined before teacher participation in a coaching model and then again following participation in at least one cycle (typically 9-15 weeks) of coaching. Student growth was measured by using student scores (as assigned to teachers) on the Northwest Education Association (NWEA) Measures of Academic Progress (MAP) test (as the median of the percentile growth scores) before participation in a structured, generalized, coaching model (looking at growth from Fall to Winter) and then again after at least one complete coaching cycle (looking at student growth from Winter to Spring). This study assumed that students typically show academic progress at a steady rate across the school year and acknowledges that some teachers may have needed more than one full cycle of coaching before meeting their individualized coaching goals. To ensure increased reliability and validity of its findings, this
study examines these data for two separate groups of teachers from two academic school years: 2016-2017 and 2017-2018.

The population for this study consisted of all K-8 teachers who were first identified as needing support by the study district, recommended for support by their principal, and who agreed to participate in a collaborative coaching model. Archived TEM and MAP data for all K-8 teachers identified as not meeting expectations in the Fall of 2016 and Fall of 2017 were requested from the district’s research and analysis department. This totaled 140 teachers for the year 2016-2017. From this original list, 53 teachers were eliminated due to being primary teachers of science, social studies, or high school subjects. 87 K-8 Reading/Math teachers from 29 individual schools remained. From this list, teachers who did not show growth as a result of coaching (of any degree) were eliminated. This left 77 teachers from 28 schools. Here, years of experience ranged from 0.5 (at the end of the study year) to 28 years; 48 of these teachers were in the first five years of teaching. For 2017-2018, data for 110 teachers was received. From this original list, 31 teachers were eliminated due to being primary teachers of science, social studies, or high school subjects. 79 K-8 Reading/Math teachers from 23 individual schools remained. Years of experience ranged from one year (at the end of the study year) to 24 years; 49 of these teachers were in their first five years of teaching. No teachers appeared in the data pool for both years, but nine schools appeared in the list for both the 2016-2017 and 2017-2018 school years.

Results

This research study explored the possible relationship between improved teacher effectiveness and student growth. In particular, it explored whether a positive change in teacher observation scores (indicating improved teacher effectiveness according to the district’s observation rubric- see Appendix C), occurring as a result of district-level generalized instructional coaching, also translates into higher levels of student growth. All teacher participants received structured, generalized (i.e. not subject specific) coaching (at least weekly...
or bi-weekly) lasting for at least one full coaching cycle (9-15 weeks). However, since the coaching provided was individualized, and a cycle is deemed complete only after a teacher meets the coaching goal, coaching cycles may have varied with some lasting up to 20 weeks. Student growth was determined using the NWEA MAP test with data from the Winter testing event (measuring student growth from Fall-Spring) and the Spring testing event (measuring student growth from Winter-Spring). Particular attention was paid to the conditional growth percentile scores using the median number. This was based on the fact that medians are less likely to be skewed by an unusually high or unusually low score (NWEA, 2017) making the median scores usually a better indicator of where the middle of the class is achieving, especially for smaller sample sizes (such as the size of a typical classroom). For the purposes of this study, the population of students who produced the scores assigned to participating teachers would have remained relatively stable (assuming some transience of students). This allowed for increased reliability and validity of findings since it helps control for extraneous student factors. This means that any significant changes in student growth determined as a result of this study can be attributed to factors other than the individual student, which includes improved teacher effectiveness following coaching.

**Research Question 1: Hypotheses and Analyses**

The first research question addressed in this study was: is there a significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores of teachers after participation in a structured coaching model?

The following hypotheses were generated and tested:

- **Hypothesis 1:**

There is a statistically significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean
MAP growth scores after participation in a structured coaching model.

- **Null Hypothesis 1:**

  There is a not a statistically significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores after participation in a structured coaching model.

To test this question, the median of all student conditional growth scores assigned to each participating teacher was pulled from the Winter MAP testing event (showing growth from the Fall testing event) as the “before” coaching sample, and then again from the Spring MAP testing event as the “after” coaching sample. The observed difference in the mean of these scores is shown in Table 4.1.

Table 4.1

<table>
<thead>
<tr>
<th>Year Group</th>
<th>Mean scores before coaching</th>
<th>Mean scores after coaching</th>
<th>Mean math scores before coaching</th>
<th>Mean math scores after coaching</th>
<th>Mean reading scores before coaching</th>
<th>Mean reading scores after coaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-2017</td>
<td>33.34</td>
<td>51.55</td>
<td>33.30</td>
<td>51.37</td>
<td>31.37</td>
<td>51.76</td>
</tr>
<tr>
<td>(n=77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-2018</td>
<td>37.98</td>
<td>52.82</td>
<td>39.14</td>
<td>54.64</td>
<td>36.76</td>
<td>50.53</td>
</tr>
<tr>
<td>(n=79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In all cases, for all student scores associated with participating teachers, and with separated out math and reading test scores (for teachers that taught either math or reading or both), an increase in student growth scores was seen in 94 percent of student scores from the 2016 group and in 86 percent of student scores from the 2017 group. To determine if this growth was statistically significant, a $t$ test was performed.

For the 2017-2018 school year, the two tailed $t$ test (based on a research question that does not indicate direction) on the difference between student scores before and after coaching for all
subjects combined determined the difference as significant, \( p < 0.01 \) (Table 4.2).

Table 4.2

Statistical t-test Analysis Comparing Student Scores for all Subjects Before and After Coaching for the 2017-2018 Year

<table>
<thead>
<tr>
<th>Value</th>
<th>Variable 1</th>
<th>Variable 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>37.98734177</td>
<td>52.8205316</td>
</tr>
<tr>
<td>Variance</td>
<td>389.1665044</td>
<td>474.2857384</td>
</tr>
<tr>
<td>Observations</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>Df</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>t Stat</td>
<td>-5.80577451</td>
<td></td>
</tr>
<tr>
<td>P(T\leq t) two-tail</td>
<td>p &lt; 0.001</td>
<td></td>
</tr>
</tbody>
</table>

As a result, since this value is less than 0.05 (the level of significance), the null hypothesis was rejected.

Additional analysis of the data looked at the growth of math and reading test scores for participating teachers separately (which included teachers that taught either math only or reading only or both math and reading). This was based on the fact that, even though all teachers received non-subject specific coaching, and the majority of participating teachers taught both math and reading (30 of the teachers from the 2017 group and 28 from the 2016 group), there were some who taught only reading (9 in the 2016 group and 8 in the 2017 group) or only math (11 in the 2016 group and 12 in the 2017 group). For 2017-2018, t-test analysis on the difference between the mean of the median student growth scores in math before coaching and after coaching was similarly significant, \( p < 0.01 \), while the t-test analysis on the difference between the mean of the median student growth scores in reading before coaching and after coaching was (slightly less) significant at \( p = 0.00078 \). This shows that there is a statistically significant difference in the before/after coaching math scores and reading scores when subjects are considered separately, and further emphasizes the significant findings found when examining all student scores together.

As with 2017-2018, analysis of the data for the 2016-2017 group on the difference
between student scores before and after coaching for all subjects combined was also significant, \( p < 0.01 \) (Table 4.3). Since this is similarly less than 0.05 (the assigned level of significance), the null hypothesis was rejected.

Table 4.3

*Statistical t-test Analysis Comparing Student Scores Before and After Coaching for the 2016-2017 Year*

<table>
<thead>
<tr>
<th>Value</th>
<th>Variable 1</th>
<th>Variable 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>33.3421053</td>
<td>51.5592105</td>
</tr>
<tr>
<td>Variance</td>
<td>218.814737</td>
<td>246.373114</td>
</tr>
<tr>
<td>Observations</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Df</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>-7.3833084</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>( p &lt; 0.001 )</td>
<td></td>
</tr>
</tbody>
</table>

Similarly, for 2016-2017, \( t \)-test analysis on the difference between student growth scores for just one subject (assigned to teachers who teach either one subject or both math and reading) before and after teachers participated in coaching was also found to be significant under both subject conditions (\( p < 0.01 \) with reading scores and \( p < 0.01 \) with math scores).

**Research Question 2: Hypotheses and Analyses**

The second research question addressed in this study was: is the degree of teacher growth in observation score (as a result of participation in a coaching model) correlated to the degree of student growth on the MAP test?

The following hypotheses were generated and tested:

- **Hypothesis 2:**
  
  There is a significant positive correlation between the degree of growth in teacher observation scores and the degree of student growth on the NWEA Map test.

- **Null Hypothesis 2:**
  
  There is no significant positive correlation between the degree of growth in teacher observation scores and the degree of student growth on the NWEA Map test.
Data analysis included a comparison of the change in teacher observation scores before and after teacher participation in generalized coaching with the change in student growth scores before and after the teacher had participated in a coaching model. Specifically, this question aimed to explore whether the degree of change in measured teacher effectiveness was correlated to the degree of change seen in student scores.

For the school year 2017-2018, the correlation between teacher growth and student growth (in both math and reading), was calculated using a Pearson correlation test little to no correlation between the degree of change in teacher effectiveness and the degree of change in student growth scores was found, \( r = 0.073 \). As a result, the null hypothesis was accepted. Non-significant correlations were also seen when these data were desegregated to examine the degree of change for math and reading student growth scores separately, \( r = -0.09 \) (with math scores only) and \( r = 0.29 \) (with reading scores only). A similarly insignificant correlation with math and reading scores grouped together was seen in data from the 2016-2017 school year, \( r = -0.18 \) for all teachers, \( r = -0.34 \) between teacher growth and student growth in reading, and \( r = 0.03 \) between teacher growth and student growth in math. This again suggests that the null hypothesis should be accepted. This is despite the fact that changes in teacher effectiveness score and in student growth scores were observed (Table 4.4).

Table 4.4

<table>
<thead>
<tr>
<th>Year Group</th>
<th>Average Change in Teacher Effectiveness scores (rubric points)</th>
<th>Average Change in student growth scores (percentile points)</th>
<th>Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-2017</td>
<td>0.6018</td>
<td>18.881</td>
<td>( r = -0.18 )</td>
</tr>
<tr>
<td>( (n = 77) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-2018</td>
<td>0.6459</td>
<td>15.585</td>
<td>( r = 0.073 )</td>
</tr>
<tr>
<td>( (n = 79) )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Indeed, despite observed changes in both teacher effectiveness and student growth scores after
coaching, in both study years, a Pearson correlation reveals that the degree of change in teacher effectiveness is not correlated with the degree of change in the student growth scores.

**Additional Discoveries**

Overall, statistical analysis of the research study questions and hypothesis revealed that

The null hypothesis should be rejected for research question 1 and accepted for research question 2. However, two additional findings of interest were discovered through the analysis process.

**Coaching and Teacher Effectiveness**

The first discovery examines the significance of changes in teacher effectiveness scores as a result of coaching. Overall, as discussed earlier, qualitative and anecdotal support for the positive impact of teacher coaching on teacher practice, though well known, has ultimately not been well supported with empirical evidence (Coburn & Russell, 2008; Cohen & Hill, 2001; Garet, et al., 2001; Supovitz & Turner, 2000). This is likely because it can be difficult to separate out factors when it comes to measuring teacher outcomes. In fact, only one key study involving a review of 44 causal studies by Kraft, et al. (2016), found instructional coaching to have a large and positive effect on teachers’ instructional practice.

The present study examined the possible relationship between improved teacher effectiveness (as a result of teacher participation in a structured, generalized coaching model) and student growth scores, but also noted the changes in teacher observation scores as part of this analysis. Table 4.4 outlines the mean growth observed in teacher observation scores as a result of district-level generalized coaching for the study years, the variance of which is seen in Figure 4.2 (for 2017-2018) and Figure 4.3 (for 2016-2017).
Even though this study did not include data from teachers who did not show growth, a $t$-test analysis was run to determine if the growth that was observed as a result of coaching was significant. Here, a $t$-test analysis of the pre and post coaching observation scores assigned to participating teachers indicate that the change in scores was significant, $p < 0.01$ for both year groups. This idea is reinforced further in Figure 4.2 and Figure 4.3, which indicate that, while the average teacher growth seems small, growth for several teachers was substantial. Indeed, the range of growth in 2016 was between 0.03 rubric points and 1.78 rubric points in the 2016 group,
and between 0.007 rubric points and 2.01 rubric points in the 2017 group. This analysis assumes the following research question: is there a significant difference in the mean observation scores of teachers before participating in a structured coaching model and the mean observation scores after participation in a structured coaching model? For this question, the null hypothesis would be rejected, and it determined that generalized coaching does have a significant impact on teacher observation score. Notably, this was true for all participating teachers grouped together as well as with novice and experienced teachers separated out. This attention to teacher experience helped inform the second discovery.

**Coaching, Teacher Experience, and Student Growth**

The second discovery addresses the issue of teacher experience as a possible factor in student achievement. The most influential studies on teacher effects on student achievement come from Rivkin et. al (2005) who, using a value-added gains model for students in grade 3-6, found that teacher quality, which is related to teacher experience, had a large impact on student achievement. Similarly, Clotfelter, et al. (2007) found that teacher experience, education, and licensure test scores all have positive effects on student achievement.

In the present study, data were pulled from two groups of teachers: 79 K-8 Reading/Math teachers from 23 individual schools using data from 2017-2018, and 77 K-8 Reading/Math teachers from 28 individual schools in 2016-2017. Here, even though years of experience ranged from 0.5-28 years, 49 teachers in the 2017 group and 48 in the 2016 group were in their first five years of teaching (40 percent of teachers from the 2017-2018 group and 62 percent of teachers in the 2016-2017 group). Given research on the role of teacher experience on both the effectiveness of instructional coaching and on student achievement, the data were desegregated further, and a separate t test run on data from teachers with 0-5 years of teaching experience only.

For both groups (2016-2017 and 2017-2018), statistical analysis revealed that the
difference in the mean of student growth scores assigned to novice teachers participating in structured, generalized coaching before and after participating in at least one cycle of coaching was indeed significant under all conditions; \( p < 0.01 \) for both year groups when looking at all subjects together, and \( p = 0.0009 \) (2016-2017) and \( p = 0.0007 \) (2017-2018) when examining student growth in math, and \( p = 0.0005 \) (2016-2017) and \( p = 0.012 \) (2017-2018) when examining student growth in reading. This would assume the following research questions: is there a significant difference between the mean student growth scores of novice (0-5 years) teachers before participation in a structured, generalized coaching model and the mean student growth scores of novice teachers after participating in a structured, generalized coaching model? Research findings suggest that, in both cases, the null hypothesis would be rejected, and it concluded that for novice K-8 teachers participating in structured, generalized coaching, there is a significant difference in the mean student growth scores before and after coaching. This finding was made significant by the fact that a comparative analysis of the data for non-novice teachers (6+) showed that the change in student scores from teachers after participation in a generalized coaching model was not significant, \( p = 0.068 \) for 2017 and \( p = 0.060 \) for 2016.

**Summary**

The findings from this study began with an analysis aligned to the following research question: is there a significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores of teachers after participation in a structured coaching model? Here a \( t \)-test analysis examined student MAP scores from the Winter testing event (assigned as before coaching) in relationship to the student MAP scores from the Spring testing event (assigned as after coaching). Grouping all teachers and all student scores (for both math and reading) together, the difference between the scores before and the scores after was revealed as statistically significant. As a result, the null hypothesis was rejected, and it concluded that there
is a statistically significant difference between the mean MAP growth scores of low
performing teachers before teacher participation in a structured coaching model and the mean
MAP growth scores after participating in a structured coaching model. This was true for data
from both the 2016-2017 school year and the 2017-2018 school year. As an additional point of
interest, additional tests were run to see if there was a significant difference between the mean
MAP scores of teachers before and after coaching when accounting for math and reading scores
separately. This was based on the fact that, even though all teachers received non-subject specific
coaching, and the majority of participating teachers taught both math and reading (30 of the
teachers from the 2017-2018 group and 28 teachers from the 2016-2017 group), there were some
who taught only reading (9 in the 2016 group and 8 in the 2017 group) or only math (11 in the
2016 group and 12 in the 2017 group). These results showed similar significance.

The second question addressed in this study examined whether the degree of teacher
growth in observation score (as a result of participation in a coaching model) was correlated to
the degree of student growth on the MAP test. For this, a Pearson correlation test was completed
and little to no correlation between these two variables determined. As a result, the null
hypothesis was accepted. Analyses included looking at math and reading scores both together
and separately, as with question one. Little to no correlation was found in all test scenarios.

Notably, this study made two additional determinations as a result of a deep analysis of
the data. Indeed, in examining teacher effectiveness measurement scores (as determined by
administrator observations using the district teacher evaluation rubric) before and after teacher
participation in a structured coaching model, a t-test analysis showed that, in addition to seeing
an observed growth in all participating teachers, that there is also a statistically significant
difference in these pre/post coaching scores. In addition, in an attempt to further desegregate data
according to factors known to affect student growth and achievement, this study took a closer
look at teachers designated as “novice” (with 0-5 years of experience) and “experienced.” Here,
in examining student growth scores before (from the Winter testing event) and after (from the Spring testing event) teacher participation in a structured, generalized coaching model, a $t$-test analysis found that there was a statistically significant difference in student scores for both groups of novice teachers (2016 and 2017) but not for either group of experienced teachers. This discovery has additional implications and offers opportunity for deeper study. These implications and opportunities will be discussed further in Chapter 5.
CHAPTER 5
Conclusions, Implications and Recommendations

Introduction

Over the last 15 years, likely in response to the increased demands placed on school districts to improve teacher practice and hold teachers accountable for student learning, the popularity of job embedded professional development programs like instructional coaching has more than doubled (Domina, Lewis, Agarwal, & Hanselman, 2015). In particular, coaching aims to improve teacher practice through job-embedded support and includes modeling, feedback, and guided implementation of research-based instructional strategies, typically provided by an identified “expert” or “master teacher” who serves as the “coach.” In almost all cases, the type of instructional coaching provided (whether school-based or district-based) falls into one of categories: subject/content specific coaching (usually literacy or math based) or generalized coaching (with emphasis on improving instructional techniques). Given the financial investment required for the implementation of both school-based and district-wide coaching programs, discussions regarding the impact of these programs on both teacher practice and student achievement are widespread. However, while there is growing evidence to support the idea that instructional coaching leads to improved teacher outcomes (Annenberg, 2004; Lockwood, et al., 2010), there is only limited research to show that it impacts student outcomes (Bright & Hensley, 2010; Kraft, et al., 2016).

The purpose of this study was to explore further the impact of instructional coaching on student achievement, with particular focus on the relationship between improved teacher effectiveness (occurring after participating in a generalized coaching model) and student growth in math and/or reading in K-8 schools in an urban school district in the Southern United States. The study sample for this research was predetermined in that all teacher participants had opted to receive weekly or bi-weekly generalized, district-level coaching in addition to whatever support was already being provided by their school’s instructional leadership team, and so appeared on
the list of teachers receiving support that was requested from the district’s research department. Teacher Effectiveness here was determined by the score (1-5) the teacher received on the district observation rubric (TEM rubric) before receiving any coaching (the summative score of all observations completed the year before receiving coaching, or from the first informal or formal observation completed in the study year) compared with the average of all scores the teacher received while being coached, and following completion of at least one full coaching cycle. The focus of the study was to determine if participation in this coaching model (resulting in an increase in observation score), led to improved student outcomes (with emphasis on student growth rather than student achievement). To determine student growth, the median of student growth scores from the NWEA MAP test (which the study district uses to assess growth of K-8 students in both math and reading) assigned to participating teachers from the Winter testing event (measuring growth from Fall to Winter, and occurring before the first coaching cycle would have started or been completed) was compared with the scores from the Spring Testing event (measuring comparative growth from Winter to Spring, and occurring after at least one full coaching cycle). Data from two groups of teachers, one group from 2016-2017 and one group from 2017-2018, were used due to allow for increased reliability and validity of the study findings.

Analysis of the received data involved a t-test analysis to determine the significance of the difference between student scores before the teacher had participated in a coaching model and student scores after the teacher had participated in a coaching model (analysis of means). For this, student conditional growth scores (the mean of the median numbers for each teacher) in math and reading were both combined and considered separately. In addition, a Pearson correlation analysis was performed to determine if the degree of change in a teacher’s observation score (following teacher participation in coaching) was correlated to the degree of change (non-directional) in the student’s growth score. This chapter includes a summary of the
results of this analysis, presents both the conclusions and implications of the associated findings, and provides recommendations for future research.

**Discussion of Results**

**Question 1: Summary and Conclusions**

The first research question addressed in this study was: is there a significant difference between the mean MAP growth scores of low performing teachers before participation in a structured coaching model and the mean MAP growth scores of teachers after participation in a structured coaching model? A discussion of the results associated with this question follow.

**Summary of Results.** To determine if any observed growth between student scores (both math and reading) obtained before the teacher had participated in a structured, generalized coaching model, and those obtained after the teacher had participated in structured, generalized coaching, was statistically significant, a $t$ test was performed. For the 2017-2018 school year, the two tailed $t$ test (based on a research question that does not indicate direction) on the difference between student scores before and after coaching for all subjects combined was found to be significant, $p < 0.01$ (Table 4.2). For the 2016-2017 school year the two tailed $t$ test on the difference between student scores before and after coaching also showed that the difference between student growth before and after coaching is significant, $p < 0.01$ (Table 4.3).

Additional analysis of the data for both years also examined the change in student scores in math and reading separately and included $t$-test analyses on a sample that included all reading scores and then on a sample that included all math scores (assigned to teachers who taught either only one subject or both math and reading) For both years, these separate $t$-test analyses found that the difference between mean student growth scores assigned to teachers who teach either one or two subjects (looking at each subject separately) before coaching and after coaching was significant for both subjects. This further emphasized the significance found when considering all subjects together. This is shown in table 5.1.
Notably, the change in student scores in reading for the 2017 group, though still statistically significant, was shown to be less significant than with math for the 2017, and less significant than both subjects for the 2016 group. This could be due to the fact that, in the study district, a new reading curriculum was implemented during the 2017 school year.

**Conclusions.** In the raw data, 94 percent of student scores from the 2016 group and 86 percent of student scores from the 2017 group showed improved growth between testing events with the average improved growth being 15 percentile points (accounting for both positive and negative growth) in 2017 and 16.6 percentile points for 2016. In support of this, the results of this study reveal that changes in student growth scores for teachers who had received at least one cycle of generalized, structured (meaning that it adheres to a formalized coaching cycle) coaching are significant, both when considering all participating teachers together and when considering math and reading scores separately. This suggests that teacher participation in coaching, where an observed growth was seen, led to a statistically significant improvement in student growth scores. The fact that students typically progress at a steady rate, and that this study used the median percentile score, and that student growth scores should be steady across all terms, means that the observed change in the mean growth scores would not have been observed without intervention. However, the fact that teaching involves so many intersecting factors, it is possible that factors other than coaching contributed to this significance. This might include teacher-level factors, and/or school-level factors. Student-factors are, however, an
unlikely a factor in determinations made by this study due to the fact that student factors were controlled for on multiple levels including: using student growth instead of student achievement, using the median growth score rather than the mean growth score, and using data from schools that have similar demographics and share Title 1 status.

Overall, these findings show that participation in generalized coaching, which results in an improvement in teacher effectiveness (as measured by improved teacher observation scores) does lead to improved student growth. This is supported by research by Kraft et al. (2016) who similarly found a relationship between student achievement and generalized coaching, even though it was weak. Kraft et al. (2016) do however also call attention to the fact that findings from studies attempting to link teacher coaching with student outcomes (whether significant or not) can be difficult to generalize due to the fact that the conditions of coaching vary so greatly across studies, and because it is difficult to control for confounding variables, including school climate (Cohen, et al., 2009), and school leadership (Leithwood, et al., 2004). Indeed, it is unknown what school-led initiatives, protocols, or supports were provided to teachers and/or students in this study, outside of the measured district-level coaching, between testing events. Finally, given that a large percentage of participating teachers were novice teachers, and the additional findings determined by this study, two additional key teacher-level factors, experience and content knowledge, likely intersected with the study results.

**Question 2: Summary and Conclusions**

The second research question addressed in this study was: is the degree of teacher growth in observation score (as a result of participation in a coaching model) correlated to the degree of student growth on the MAP test? A discussion of the results associated with this question follow.

**Summary.** Data analysis for Question 2 included a comparison of teacher observation scores before and after teacher participation in generalized coaching (looking at the change in score) with student growth scores before and after the teacher had participated in a coaching
model. Specifically, this question aimed to explore whether the degree of change in measured teacher effectiveness was correlated to the degree of change seen in student scores.

For the school year 2017-2018, the correlation between teacher growth (following participation in a coaching program) and student growth (in both math and reading), was calculated using a Pearson correlation test and little to no correlation between the degree of change in teacher effectiveness and the degree of change in student growth scores was found, \( r = 0.073 \). As a result, the null hypothesis was accepted. Similar non-significant correlations were seen when the data were segregated to examine the degree of change for math and reading student growth scores separately, \( r = -0.09 \) (with math scores only) and \( r = 0.29 \) (with reading scores only). Insignificant correlations were also seen in data from the 2016-2017 school year, \( r = -0.18 \) for all teachers, with math and reading grouped together, \( r = -0.34 \) between teacher growth and student growth in reading, and \( r = 0.03 \) between teacher growth and student growth in math. Thus, despite observed changes in both teacher effectiveness and student growth scores after coaching, in both study years, a Pearson correlation reveals that the degree of change in teacher effectiveness is not correlated with the degree of change in the student growth scores.

**Conclusions.** The raw teacher data from this study revealed an average increase in teacher observation score of 0.60 rubric points in the 2016 group, and an average growth of 0.64 rubric points for the 2017 group. Similarly, in the raw student data, 94 percent of student scores from the 2016 group an 86 percent of student scores from the 2017 group showed changes in growth, with an average growth change of 15 percentile points (accounting for both positive and negative growth) in 2017 and 16.60 percentile points for 2016. A Pearson correlation analysis of these changes, however, shows that degree of change in teacher practice is not correlated to the degree in change in student growth scores. Either way, even though the degree of teacher growth was not found to be correlated with the degree of change in student growth, based on the analysis of Question 1, this growth, both in teacher score and student score was still significant. In
addition, the fact that the raw data shows that mean student scores after coaching moved to the 50th percentile and above is critical. This is shown in Table 4.1 (reproduced here below).

<table>
<thead>
<tr>
<th>Year Group</th>
<th>Mean scores before coaching</th>
<th>Mean scores after coaching</th>
<th>Mean math scores before coaching</th>
<th>Mean math scores after coaching</th>
<th>Mean reading scores before coaching</th>
<th>Mean reading scores after coaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-2017</td>
<td>33.34</td>
<td>51.55</td>
<td>33.30</td>
<td>51.37</td>
<td>31.37</td>
<td>51.76</td>
</tr>
<tr>
<td>(n=77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-2018</td>
<td>37.98</td>
<td>52.82</td>
<td>39.14</td>
<td>54.64</td>
<td>36.76</td>
<td>50.53</td>
</tr>
<tr>
<td>(n=79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An NWEA MAP growth score at the 50th percentile is the threshold at which a student has met their growth goal (and indicates that a student’s growth is greater than 50% of all students in the NWEA norm group). As a result, the median growth score expectation for all teachers in all terms, according to NWEA, is 50. 18 percent of teachers in the 2016 group had a median growth percentile that met this expectation before coaching, compared to 49 percent after coaching, while 25 percent of teachers in the 2017 group had a median growth percentile that met this expectation in the first testing event (before coaching), compared to 51 percent in the second testing event (after coaching). This means that the number of students reaching their goal (which would have been different for each event and determined by their past performance) doubled between testing events. This is an indicator of improved teacher effectiveness.

**Additional Findings: Summary and Conclusions**

This study revealed two additional findings of interest through the analysis process: an significant difference between teacher observation scores before participation in a coaching model and after, and an insignificant difference between the mean student scores of experienced, compared to novice teachers before and after teacher participation in a structured, generalized coaching model. These findings are discussed further below.

**Coaching and Teacher Effectiveness.** The first discovery examines the significance of changes in teacher effectiveness scores following teacher participation in coaching. The present
study examined the possible relationship between improved teacher effectiveness (following participation in a structured, generalized coaching model) and student growth scores. Since participants for the study were selected based on their participation in a coaching model, and their data excluded if they did not show growth in teacher observation score, this study assumes that any growth seen in teacher score was the result of varying degrees of coaching. In particular, this study found that teacher participants had an average growth of 0.64 rubric points in the 2017 group and 0.60 rubric points in the 2016 group. A t-test analysis of these pre and post coaching observation scores assigned to teachers indicate that the change in scores was significant, where \( p < 0.01 \).

This finding adds to a body of research that attempts to link teacher coaching with teacher practice. Indeed, even though qualitative and anecdotal support for the positive impact of teacher coaching on teacher practice is well known, it has ultimately not been well supported with empirical evidence (Coburn & Russell, 2008; Cohen & Hill, 2001; Garet, et al., 2001; Supovitz & Turner, 2000). This is likely because it can be difficult to separate out factors when it comes to measuring teacher outcomes. In fact, only one key study involving a review of 44 causal studies by Kraft, Blazar & Hogan (2016), found instructional coaching to have a large and positive effect on teachers’ instructional practice. That said, given that there are a number of extraneous variables that cannot be determined and accounted for, including whether the observations were conducted by the same or different administrator, whether the observation by the administrator matches the observations of the instructional coach (who arguably would have “observed” the teacher more frequently), and the number of observations included in both the “before” coaching and “after” coaching score, it is possible that other factors intersected with coaching in providing this significant result. Notably, 88 percent of the teachers in the 2017 group moved from below the district’s threshold for effectiveness (2.75/5) to above the threshold, while 68 percent of teachers in the 2016 group moved from below the district’s
threshold for effectiveness (2.75/5) to above the threshold following participation in the coaching model. This suggests that teacher growth as a result of coaching was both statistically significant and impactful. This is especially so given the fact that the study district, like many school districts across the country, uses teacher observation scores as a measure of effectiveness to determine continued employment. In particular, teachers with low observation scores (below 2.75) that persist across multiple years, despite coaching efforts (internal and/or external), are subject to non-renewal or dismissal.

**Teacher Experience and Student Growth.** The second discovery addresses the issue of teacher experience as a possible factor in student achievement. Given research on the role of teacher experience on both the effectiveness of instructional coaching and on student achievement, as part of analyses for question 1, this study also conducted t-test analyses of the data by looking at data for novice teachers (0-5 years) and experienced (6+ years) teachers separately. For both groups (2016-2017 and 2017-2018), statistical analyses revealed that the difference in the mean of student growth scores assigned to novice teachers participating in structured, generalized coaching before and after participating in at least one cycle of coaching was also significant, \( p < 0.01 \). This was expected since a large number of participating teachers were novice teachers, and the t-test analysis of the whole group revealed that the difference between before coaching and after coaching growth scores for all teachers combined was also found to be significant, \( p < 0.01 \). Further analysis of the data assigned to experienced teachers (defined here as having six or more years of experience) showed that, contrastingly, the difference in the mean of student growth scores assigned to experienced teachers participating in structured, generalized coaching before and after participating in at least one cycle of coaching was not significant, \( p = 0.068 \) for the 2017 group and \( p = 0.06 \) for the 2016 group.

This finding is well supported by research. Indeed, Gamm (2013) in a study that examined the differences in coaching novice (defined as 0-5 years of experience) versus
experienced teachers (defined as 6+), found that experienced teachers identified for coaching tended to be more resistant to change, less likely to implement recommended strategies, less reflective about their practice, and had less coaching interactions than novice teachers. As a result, coaches observed less growth in their experienced teachers than they did in their novice teachers. These factors could help explain why student scores before/after teacher coaching were not significant when assigned to experienced teachers versus novice teachers. However, interestingly, despite the fact that the growth in student scores assigned to experienced teachers was not significant, the change in teacher observation score following coaching participation remained significant for all groups. One explanation for this is that experienced teachers are better able to navigate the politics and procedures of teacher observations (Gamm, 2013) and so are more likely to score higher on teacher observation rubrics. This suggests that these scores are more indicative of what the teacher can do as opposed to what they actually do on a daily basis. The idea that experienced teachers tend to show less growth on observation scores is supported by the fact that in the current study, where, despite showing a similarly statistically significant improvement in observation score, the raw data showed that experienced teachers grew less (0.59 rubric points in 2017 and 0.56 rubric points in 2016) than novice teachers did (0.63 rubric points in 2017 and 0.58 rubric points in 2016). School-level and coach-level factors are likely intervening variables here given that newer teachers (with 47 percent of novice in the 2017 group, and 80 percent of novice teachers in the 2016 group being in their first year of teaching) typically receive additional layers of support including that from a new teacher mentor. These teachers would have also received extensive classroom feedback from their administrative leadership team as they would have been assigned to the novice evaluation track (meaning that they would been required to have four formal observations). This building-level coaching would also have included assistance with content area planning, which means that these teachers are likely to have received more feedback and more coaching interactions than the participating
experienced teachers.

**Implications**

The basic assumptions underpinning this study are that, as more and more schools and districts are turning to instructional coaching as a means for teacher development, conversations regarding whether coaching actually improves teacher and student outcomes are in the forefront of educational policy and planning. In particular, this study explores the differences in student (conditional) growth (on NWEA MAP reading and math tests) on tests completed before the teacher participated in a structured, generalized coaching model, compared to after, looking at data from two groups of teachers, one each from two separate year groups. This study found that the changes in student growth index were significant for all participating teachers and for novice teachers when considered separately from experienced teachers. Comparatively, the growth in student scores assigned to participating experienced (with 6 or more years of experience) teachers between testing events were not found to be significant. Table 5.2 summarizes these findings.

**Table 5.2**

*A t-test Analysis of the Difference Between Student Scores Before And After Teacher Participation in a Structured Coaching Model*

<table>
<thead>
<tr>
<th>Score Type</th>
<th>p value for 2016 group</th>
<th>p value for 2017 group</th>
</tr>
</thead>
<tbody>
<tr>
<td>All student scores assigned to all participating teachers</td>
<td><em>p</em> &lt; 0.01</td>
<td><em>p</em> &lt; 0.01</td>
</tr>
<tr>
<td>All math scores assigned to teachers</td>
<td><em>p</em> &lt; 0.01</td>
<td><em>p</em> &lt; 0.01</td>
</tr>
<tr>
<td>All reading scores assigned to teachers</td>
<td><em>p</em> &lt; 0.01</td>
<td><em>p</em> &lt; 0.01</td>
</tr>
<tr>
<td>All student scores assigned to novice teachers only</td>
<td><em>p</em> &lt; 0.01</td>
<td><em>p</em> &lt; 0.01</td>
</tr>
<tr>
<td>All student scores assigned to experienced teachers only</td>
<td><em>p</em> = 0.06</td>
<td><em>p</em> = 0.06</td>
</tr>
</tbody>
</table>

The study findings, as a result of $t$-test analysis, indicate that there is a statistically significant difference between student growth scores before and after teacher participation in coaching. This means that the observed student growth between testing events (as evidenced by an increase in
the overall mean of the median growth scores between testing events), can be attributed, at least in part, to teacher participation in a structured generalized coaching model. In addition, further examination of math and reading scores separately (as assigned to teachers who teach either one subject or both) found that changes in student growth were similarly significant (Table 5.2). The fact that this finding is true under multiple conditions suggests further that coaching has a significant impact on student achievement. Given the substantial financial investment that districts put into coaching this has far reaching implications for those who turn to coaching as a way to improve student outcomes (by way of improving teacher outcomes). This is especially so given the fact that there was evidence that generalized coaching improved teacher outcomes (only teachers who showed some growth as a result of coaching were included in this study), and evidence that this growth (following coaching) was significant. This was true when including all participating teachers, and when looking at novice teachers separately from more experienced teachers. This is shown in Table 5.3.

Table 5.3

<table>
<thead>
<tr>
<th>Teacher Group</th>
<th>P value for t-test analysis from 2017 group</th>
<th>P value for t-test analysis from 2016 group</th>
</tr>
</thead>
<tbody>
<tr>
<td>All teachers</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>All novice Teachers</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>All experienced Teachers</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.01$</td>
</tr>
</tbody>
</table>

The results from the analyses of novice versus experienced teachers reiterates the idea that for participating teachers who had more than six years of experience, the change in student growth scores was not significant, while the change in student scores for novice participating teachers was. This offers an important contribution to educational research as these findings suggest that coaching has a greater impact on the student growth scores of novice teachers (albeit a similar effect on teacher growth scores) than on experienced teachers. This is supported by research that
suggests that new teacher coaching tends to lead to changes in practice that have a more wide-spread impact (affecting culture, instruction, and achievement simultaneously) compared to that seen when coaching experienced teachers, especially when that new teacher coaching is ongoing \((The\ New\ Teacher\ Center,\ 2016)\). In addition, the fact that research suggests that new teachers tend to need, and receive, different types of coaching means that instructional coaches have to interact with (and are trained as such) new teachers in a different way than they do more experienced teachers. Marzano, Frontier, and Livingston (2011) refer to this need in their research and suggest that effective teachers require proficiency in three general areas, which cannot be separated from each other: routine strategies (dealing with routines, rules, culture, etc.), content strategies, and strategies enacted on the spot (checking for understanding, enforcing rules, etc.); new teachers in this study likely received coaching support in all three areas (as opposed to experienced teachers who may have only received support in two or three), which may have contributed to this finding. The fact that there is no way to know this for sure, however, offers the potential for further study.

Finally, this study examined whether the degree of change in teacher practice (where all teachers grew to some extent) was correlated to the degree of change in scores for students (as assigned to teachers). Little to no correlation was found (Table 5.3).

<table>
<thead>
<tr>
<th>Participant Group</th>
<th>Pearson Correlation (math and reading)</th>
<th>Pearson Correlation (math)</th>
<th>Pearson Correlation (reading)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-2018</td>
<td>(r = 0.073)</td>
<td>(r = -0.097)</td>
<td>(r = 0.297)</td>
</tr>
<tr>
<td>2016-2017</td>
<td>(r = -0.182)</td>
<td>(r = 0.037)</td>
<td>(r = -0.343)</td>
</tr>
</tbody>
</table>

Given the importance assigned to teacher observation scores, and to student scores, the idea that the degree of change in one is not correlated to the degree of change in the other is significant
and may have implications for how districts measure the success of their coaching programs. Indeed, since instructional coaches tend to observe teacher practice more frequently than administrators (who complete and score formal observations using an evaluation rubric), formal observation scores, as used in this study, may not provide a complete picture of the ways in which, and degree to which, coaching improves teacher practice. This might be better provided by a narrative of the coaching interactions and of teacher practices. For the study district, this narrative is provided for all participating teachers at the end of the academic school year but is not necessarily used by schools and principals (since it resides at the district level) in determining next steps for participating teachers (who typically receive only one year of coaching). There is thus potential for this narrative to be used to create personal development plans for participating teachers and for school districts to determine trends.

Summary

Research into the effectiveness of coaching on both teacher practice and student outcomes remains limited due to the fact that it is so difficult to separate out confounding variables. Indeed, even though analyses of several coaching programs have found that coaching provides teachers with essential tools to become more effective instructors, which essentially translates into improved student achievement (Biancarosa, et al., 2010; Marsh et al., 2010; Sailors & Price, 2010), this research focuses primarily on reading instruction for the early grades. This study, in an examination of student outcomes for K-8 teachers in an urban school district in the United States, who participated in at least one (generalized) coaching cycle, contributes to this research, and found that changes in teacher observation scores did have a significant impact on student growth scores (when considering all participating teachers and when considering novice teachers apart from experienced teachers). Additionally, results from this research study found that the change in teacher observation score (before and after coaching) was significant under all conditions (all participating teachers, novice teachers, and experienced teachers), but
that the degree of change in a teacher’s score is not correlated to the degree of change in student scores.

**Recommendations for Further Study**

The results in this study revealed that there is a significant difference in student growth scores (for math and reading) before and after coaching (and as a result of improved teacher effectiveness). Here, given the likely intersection of teacher and school-level factors in this study, a study that attempts to examine which coaching practices (including pairing coaching with seminar style professional development sessions) yield greater results on both teacher outcomes and student outcomes, would have heavy implications for how districts select and train their coaches. Similarly, the fact that this study had additional findings regarding coaching for novice versus experienced teachers, offers potential for further study. In particular, an extended study (two years) to investigate the effect of coaching on student and teacher outcomes for novice teachers, as well as a study that looks more specifically at the role of teacher experience in the efficacy of coaching, would greatly extend the findings from this study. Finally, a qualitative study examining the differences in perceptions of coaching from novice teachers versus experienced teachers would further extend the findings from this study that note distinctions between the efficacy of coaching (especially on student growth scores) for these two groups.
References


doi:http://0-dx.doi.org.library.acaweb.org/10.1007/s10964-011-9652-8


Student Achievement? NBER Working Paper 12828.


Educational Improvement & Opportunity in 50 Cities. Center for Reinventing Public Education.


http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.901.1275&rep=rep1&type=pdf


https://sites.nationalacademies.org/cs/groups/dbasesite/documents/webpage/dbasse_160397.pdf


School Climate Center.


http://www.ncl.ac.uk


New Jersey Department of Education. (2010). Retrieved from
http://www.state.nj.us/education/schools/achievement/2010/njask7/summary.pdf


Northern, A. (2016). The effects of good teacher professional development on student
achievement. Thomas Fordham Institute. Retrieved from:
https://edexcellence.net/articles/the-effects-of-good-teacher-professional-development-on-student-achievement


Teaching and Teacher Education 13(3): 245–58

administrators: A national study of principal quality. National Association of Secondary

10.1093/acrefore/9780190264093.013.27

Pennsylvania Institution for Coaching (2018). What is an Instructional Coach? Retrieved from:

rates, when gauging school effectiveness. Retrieved from:


Research.


Center for Educational Policy and Management.


The University of Florida Lastinger Center for Learning, Learning Forward, & Public Impact. (2016). Coaching for impact: Six pillars to create coaching roles that achieve their potential to improve teaching and learning. Gainesville: University of Florida Lastinger


APPENDIX A
IRB Approval
IRB

Steve A. Davidson
Thu 11/29/2018 7:51 AM
To: Michelle R Hope <mrhope@cn.edu>

I received the following email from the chair of the IRB committee. You are approved to proceed! **When you obtain your district’s approval, please send me a copy!**

(His comment is a result of the email I sent reminding him of the need for approval prior to your district's assent to conduct your research.)

================================
IRB
B

Thu 11/29/2018 8:21 AM
I overlooked that – this request is approved – thanks for reminding me

In His service,

Gregory A. Casalenuovo, PhD, APRN, FNP-BC, FNP-C
Professor of Nursing
Associate Director, Honors Program
Carson-Newman University
C-N Box 71883
Jefferson City, TN 37760

Dr. Steve A. Davidson
Associate Professor, Graduate Education
CNU Box 72013 | Jefferson City, TN 37760
Office #35 | Stokely Memorial Building
Office: 865.471.3319 | Fax: 865.471.3475

CARSON-NEWMAN
A CHRISTIAN UNIVERSITY

http://outlook.office.com/owa/?viewmodel=ReadMessageItem&ItemID...wdTUCf%2BVIAIn1Fi5AAA%3D&IsPrintView=1&wid=32&ispopout=1&path=
APPENDIX B
District Approval
To: Michelle Hope  
Re: Research Proposal

After consideration of your proposal, "The Effect of Improved Teacher Effectiveness on Student Growth," we have approved your request to conduct this study in ... You should use this letter as official notification of approval for your study.

Note that the district-level approval being granted with this letter does not obligate any school or any person to participate in this project. Approval by the principal of any participating school is still needed before the study can begin at that school. Also, individuals must be given the option of not participating.

Approval also requires that you use findings only for the purpose of the study described in the proposal.

We look forward to working with you in the completion of this project.

Sincerely,

Jeffery A. Shive
APPENDIX  C
Core Components of the Study District’s Observation Rubric
## Teach 1 – Objective Driven Lessons

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Level 1 – Significantly Below Expectations</th>
<th>Level 2 – Below Expectations</th>
<th>Level 3 – Meeting Expectations</th>
<th>Level 4 – Above Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective(s) / Standards</td>
<td>The following text describes what is observed.</td>
<td>The following text describes what is observed.</td>
<td>The following text describes what is observed.</td>
<td>The following text describes what is observed.</td>
</tr>
<tr>
<td>Appropriate Language</td>
<td>1. Teacher does not communicate lesson objective(s) and excuses how the objective(s) is/are related to standards.</td>
<td>1. Teacher communicates lesson objective(s) but does not explain how the objective(s) is/are related to standards.</td>
<td>1. Teacher communicates lesson objective(s) to students in relationship to standards.</td>
<td>Level 4 – Evidence fully supporting Level 3 is present, as well as all of the following:</td>
</tr>
<tr>
<td></td>
<td>2. Teacher uses language that is incorrect and inappropriate for the context.</td>
<td>2. Teacher uses language that is not developmentally appropriate.</td>
<td>2. Teacher uses developmentally appropriate language.</td>
<td>Level 5 – Evidence fully supporting Level 3 is present, as well as all of the following:</td>
</tr>
<tr>
<td></td>
<td>Models Mastery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Teacher does not explain or model what mastery of the objective(s) or related performance tasks look like.</td>
<td>3. Teacher inappropriately explains or models what mastery of the objective(s) and/or related performance tasks look like.</td>
<td>3. Teacher explains or models what mastery of the objective(s) and/or related performance tasks look like.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Teacher provides opportunities for engagement, which are completely disconnected from the lesson objective(s) or does not provide opportunities for engagement with the objective(s).</td>
<td>4. Teacher provides limited opportunities for engagement in the lesson objective(s) and/or opportunities for engagement have minimal connection with the lesson objective(s).</td>
<td>4. Teacher provides multiple opportunities for engagement in the lesson objective(s) including connecting to prior knowledge.</td>
<td>1. Teacher provides clear explanation and/or modeling of what mastery of objective(s) and/or exemplary student work looks like.</td>
</tr>
<tr>
<td></td>
<td>Understand Objective(s)</td>
<td></td>
<td></td>
<td>2. Teacher actively and effectively engages students in the process of connecting the lesson with their prior knowledge.</td>
</tr>
<tr>
<td></td>
<td>5. Students struggle to retell/demonstrate the objective(s) or explain the tasks they are completing.</td>
<td>5. Students can retell the objective(s) or describe/demonstrate tasks; however, they are unable to make connections to what they are learning.</td>
<td>5. Most students can explain or demonstrate what they are learning beyond simply repeating the stated or posted objective(s).</td>
<td>3. Most students can explain or demonstrate the lesson objective(s) within the context of the related standard(s) and test or demonstrate the importance of their learning.</td>
</tr>
<tr>
<td></td>
<td>Importance of Learning</td>
<td></td>
<td></td>
<td>4. Students model or explain mastery to other students.</td>
</tr>
<tr>
<td></td>
<td>6. Students are unable to describe the importance of their learning.</td>
<td>6. Students offer inaccurate, neutral or demonstrate, regarding the importance of their learning.</td>
<td>6. Most students can explain the importance of their learning.</td>
<td>5. All students can describe how their learning will be assessed.</td>
</tr>
<tr>
<td></td>
<td>Assessment of Student Learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. Students are unable to describe how their learning will be assessed.</td>
<td>7. Few students can describe how their learning will be assessed.</td>
<td>7. Most students can describe how their learning will be assessed.</td>
<td></td>
</tr>
</tbody>
</table>

1. To determine if students can explain or demonstrate knowledge of the objective(s), observers should ask students questions in non-intrusive ways about the objective(s). Observers should also focus keenly on student responses and conversations to gauge their understanding of the objective(s) meaning and importance.

2. Students should see examples and non-examples, in some cases, that relate to the instructional activity to support student understanding. Examples can be from previous students’ work or teacher-created exemplars.
### Teach 2 – Explain Content

<table>
<thead>
<tr>
<th>Description</th>
<th>Level 1 – Significantly Below Expectations</th>
<th>Level 2 – Below Expectations</th>
<th>Level 3 – Meeting Expectations</th>
<th>Level 4 – Above Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Explanations</td>
<td>1. Teacher’s explanations/demonstrations of content are unclear, incoherent, or inaccurate.</td>
<td>1. Teacher’s explanations/demonstrations of content are generally clear, coherent, and accurate, with a few exceptions.</td>
<td>1. Teacher’s explanations/demonstrations of content are clear and accurate, and build student understanding of content.</td>
<td>Level 4 – Evidence fully supporting Level 3 is present, as well as one of the following:</td>
</tr>
<tr>
<td></td>
<td>2. Teacher does not build student understanding of content.</td>
<td>2. Teacher may not be entirely effective in building student understanding of content.</td>
<td>2. Teacher provides logical sequencing of essential information.</td>
<td>1. Teacher explains/demonstrates concept(s) in a way that activity involves most students in the learning process and promotes student interest in the content.</td>
</tr>
<tr>
<td></td>
<td>3. Teacher does not utilize multiple perspectives/approaches to solve problems or interpret text/content.</td>
<td>3. Teacher utilizes limited perspectives/approaches to solve problems or interpret text/content.</td>
<td>3. Teacher utilizes multiple perspectives/approaches to solve problems or interpret text/content.</td>
<td>2. Students make independent connections through classroom interactions demonstrating that they understand the content levels ranging from basics to complex.</td>
</tr>
<tr>
<td></td>
<td>4. Teacher does not make connections between content areas, students’ experiences and interests, or current events.</td>
<td>4. Teacher makes minimal/limited connections between content areas, students’ experiences and interests, or connections do not build student understanding of content.</td>
<td>4. Teacher makes relevant connections between content areas, students’ experiences and interests, or current events.</td>
<td>3. Students, when possible, consider multiple perspectives and approaches to learning.</td>
</tr>
<tr>
<td></td>
<td>5. Teacher uses explanations that are developmentally inappropriate and include academic language and definitions that are completely unclear or improper.</td>
<td>5. Teacher uses explanations that are somewhat developmentally appropriate and include academic language and definitions that are not completely clear or precise.</td>
<td>5. Teacher uses explanations that are developmentally appropriate and include academic language that is clear and precise.</td>
<td>4. Presentation of content includes modeling by the students to demonstrate performance expectations.</td>
</tr>
<tr>
<td></td>
<td>6. Teacher rigidly adheres to the initial plan for explaining content, even when it is clear that an explanation does not effectively lead students to understand the concept.</td>
<td>6. Teacher re-explains in the same way rather than providing an alternative explanation when students do not understand.</td>
<td>6. Teacher demonstrates alternative ways to explain concepts effectively.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. Students may demonstrate confusion by teacher explanations, frustration, or disengagement because of the teacher’s unclear explanations.</td>
<td>7. Students may ask some clarifying questions showing that they are confused by the teacher’s explanations.</td>
<td>7. Students may ask some clarifying questions providing information and feedback that the teacher uses to monitor and adjust instruction.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8. Presentation of content provides no modeling by the teacher and demonstrates no performance expectations that are unclear or confusing to students.</td>
<td>8. Presentation of content by teacher includes modeling by the teacher and demonstrates performance expectations.</td>
<td>8. Teacher models to demonstrate performance expectations.</td>
<td></td>
</tr>
</tbody>
</table>

*Logical sequencing of essential information is included, but is limited to, using visuals to establish the purpose of the lesson, preview the organization, and summarize its content.*

### Teach 3 – Appropriately Challenging Work

<table>
<thead>
<tr>
<th>Description</th>
<th>Level 1 – Significantly Below Expectations</th>
<th>Level 2 – Below Expectations</th>
<th>Level 3 – Meeting Expectations</th>
<th>Level 4 – Above Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaches Students</td>
<td>1. Teacher does not engage students in appropriately challenging work and does not reach at least one level of mastery.</td>
<td>1. Teacher attempts to engage students in appropriately challenging work; however, the teacher does not reach at least one level of mastery.</td>
<td>1. Teacher engages students in appropriately challenging work by reaching most students.</td>
<td>Level 4 – Evidence fully supporting Level 3 is present, as well as one of the following:</td>
</tr>
<tr>
<td></td>
<td>2. Teacher does not challenge students.</td>
<td>2. Teacher attempts to challenge students.</td>
<td>2. Teacher challenges students.</td>
<td>1. Teacher engages all students in appropriately challenging work by ensuring that the lesson includes appropriately complex texts, tasks, and activities that move students beyond their current mastery level.</td>
</tr>
<tr>
<td>Challenges Students</td>
<td>3. Teacher does not meet the needs and learning styles of students.</td>
<td>3. Teacher attends to limited learning levels/styles, not meeting the needs of students.</td>
<td>3. Teacher meets students at appropriate learning levels/styles.</td>
<td>2. Teacher ensures most students (at low, middle, and high achieving levels) move beyond current mastery levels.</td>
</tr>
<tr>
<td>Learning Levels/Styles</td>
<td>4. Teacher does not include complex texts/tasks to support students’ mastery of planned learning objective(s).</td>
<td>4. Teacher sporadically or occasionally uses appropriately complex texts/tasks to support students’ mastery of planned learning objective(s).</td>
<td>4. Teacher includes appropriately complex texts/tasks to support students’ mastery of objectives.</td>
<td>3. Students are engaged in accountable talk with their peers, as appropriate.</td>
</tr>
<tr>
<td>Complex Texts/Tasks</td>
<td>5. Teacher does not incorporate activities and materials that sustain student attention at learning levels/styles.</td>
<td>5. Teacher incorporates activities and materials that sustain student attention at learning levels/styles at certain points in the lesson.</td>
<td>5. Teacher incorporates activities and materials that sustain student attention at appropriate learning levels/styles throughout the lesson.</td>
<td>4. Teacher designs the lesson to incorporate resources that extend beyond the district’s curriculum.</td>
</tr>
</tbody>
</table>

*To make content reachable for students, a teacher might differentiate content, process, or product (using strategies that might include flexible grouping, chunking grade appropriate tasks, or tiered assignments) in order to ensure that students are able to access the lesson so that they eventually meet or exceed grade level standards.*

*In order for strategies to lead students to a deeper understanding of the content, a teacher must understand students’ current levels of performance and then purposefully design instructional strategies that will scaffold student learning to a deeper level so that they meet or exceed grade level expectations. Scaffold is defined as “breaking tasks down into smaller elements.” Examples of scaffolding include: activating prior knowledge, breaking tasks into smaller parts, and modeling or having student verbalize their thinking process. There are many effective techniques for scaffolding and different levels should be used always. Conversely, the techniques can also be used ineffectively. In order to be effective, the scaffolding techniques must be well-executed and appropriate to the objectives, and thus succeed in addressing the students’ misunderstandings.*

*Learning styles include auditory, visual, and kinesthetic (tactile).*

*Scooter’s instructional maps guide test text selections. Task complexity refers to engaging students in ways that align to Bloom’s levels of rigor regarding the lesson’s standards-based objective(s). A task whose standard requires the rigor level of “evaluation” should not be limited to activities that only require “remembering” or “applying.”*
## Teach 4 – Content Engagement

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Level 1 – Significantly Below Expectations</th>
<th>Level 2 – Below Expectations</th>
<th>Level 3 – Meeting Expectations</th>
<th>Level 4 – Above Expectations</th>
<th>Level 5 – Significantly Above Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement Strategies</td>
<td>1. Teacher’s engagement strategies are not aligned to the lesson objective(s) and do not have a clear, intentional purpose.</td>
<td>1. Teacher’s engagement strategies are aligned to the lesson objective(s) and have a purpose relative to accomplishing the objective(s) while others keep students busy without a purposeful use of time.</td>
<td>1. Teacher’s engagement strategies are aligned to the lesson objective(s) and have a purpose consistent with the provided objectives.</td>
<td>1. Evidence fully supporting Level 3 is present, as well as one of the following:</td>
<td>1. Teacher’s engagement strategies provide all students with choice, as appropriate.</td>
</tr>
<tr>
<td>Teacher / Student Balance</td>
<td>2. Lesson has an imbalance of teacher-directed instruction and student-centered learning.</td>
<td>2. Lesson is teacher-directed with student-centered learning happening sporadically and with little connection to the lesson.</td>
<td>2. Lesson has a balance of teacher-directed instruction and student-centered learning.</td>
<td>2. Students know how to self-select strategies that will help them master lesson objective(s).</td>
<td>2. Students know how to self-select strategies that will help them master lesson objective(s).</td>
</tr>
<tr>
<td>Differentiation / Scaffolding</td>
<td>3. Teacher’s strategies inhibit students from meeting lesson objective(s), e.g., using inappropriate scaffolding and differentiation.</td>
<td>3. Teacher’s strategies allow few students to meet lesson objective(s) through appropriate scaffolding and differentiation.</td>
<td>3. Teacher’s strategies enable students to meet lesson objective(s) with appropriate scaffolding and differentiation.</td>
<td>3. Teacher shows students to explain or demonstrate the strategies they use and how the strategies relate to what they are learning in terms of content standards.</td>
<td>3. Teacher shows students to explain or demonstrate the strategies they use and how the strategies relate to what they are learning in terms of content standards.</td>
</tr>
<tr>
<td>Student Practice</td>
<td>4. Teacher does not allow students to practice, apply and demonstrate content mastery through discussion and/or writing about complex texts, tasks, or concepts.</td>
<td>4. Teacher’s model allows students to practice, apply and demonstrate what they are learning through discussion and/or writing about complex texts, tasks, or concepts.</td>
<td>4. Teacher allows students to practice, apply and demonstrate content mastery through discussion and/or writing about complex texts, tasks, or concepts.</td>
<td>4. Teacher shows students’ knowledge of the strategies they use and how the strategies relate to what they are learning in terms of content standards.</td>
<td>4. Teacher shows students’ knowledge of the strategies they use and how the strategies relate to what they are learning in terms of content standards.</td>
</tr>
<tr>
<td>Problem Solving Strategies</td>
<td>5. Teacher either does not teach problem-solving strategies, or the teacher implements strategies that are not clearly related to the learning objective(s), content, or activity.</td>
<td>5. Teacher attempts to implement problem-solving strategies without effectively modeling or engaging students in the process, and/or students struggle to implement strategies without direct instruction due to limited teacher support.</td>
<td>5. Teacher models and implements appropriate strategies that teach or reinforce one or more of the following problem-solving types:</td>
<td>5. Teacher explicitly teaches one or more of the following problem-solving types:</td>
<td>5. Teacher explicitly teaches one or more of the following problem-solving types:</td>
</tr>
</tbody>
</table>

---

1. To determine if students can explain or demonstrate knowledge of the objective(s), observers should ask students questions in non-intrusive ways about the objective(s). Observers should also focus keenly on student responses and conversations to gauge their understanding of the objective’s meaning and importance.

2. To make observations for students, a teacher might differentiate content, process, or product (using strategies that might include flexible grouping, chunking grade-appropriate texts, or tiered assignments) in order to ensure that students are able to access the lesson so that they eventually meet or exceed grade level standards.
## Teach 5 – Higher-Level Thinking Skills

<table>
<thead>
<tr>
<th>Description</th>
<th>Level 1 – Significantly Below Expectations</th>
<th>Level 2 – Below Expectations</th>
<th>Level 3 – Meeting Expectations</th>
<th>Level 4 – Above Expectations</th>
<th>Level 5 – Significantly Above Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop Higher-Level Thinking Skills</td>
<td>1. Teacher attempts to improve the student’s critical thinking skills, but does not engage students in activities, tasks, and/or discussions that build on a solid foundation of knowledge leading to higher-order thinking skills.</td>
<td>1. Teacher attempts to ensure the lesson develops higher-level thinking skills by engaging students in tasks, activities, and discussions that build on a solid foundation of knowledge. Teachers often fail to demand stringent thinking or require students to use higher-order thinking skills.</td>
<td>1. Teacher ensures the lesson develops higher-level thinking skills by engaging students in activities, tasks, and/or discussions that build on a solid foundation of knowledge, and only requires students to use higher-order thinking skills.</td>
<td>1. Evidence fully supporting level 5 is present, as well as one of the following: 1. Teacher ensures the lesson develops higher-level thinking skills by challenging all students to engage with complex materials and think critically, and 2. Teacher actively engages students in higher-level thinking.</td>
<td>1. Evidence fully supporting level 5 is present, as well as one of the following: 1. Teacher ensures the lesson develops higher-level thinking skills by challenging all students to engage with complex materials and think critically, and 2. Teacher actively engages students in higher-level thinking.</td>
</tr>
</tbody>
</table>

### Thought Process

<table>
<thead>
<tr>
<th>1. Teacher does not model the thought process for generating and asking questions, so that students begin to generate their own questions.</th>
<th>1. Teacher models his or her thought process for generating and asking questions, but does not ask students to develop their own questions.</th>
<th>1. Teacher models his or her thought process for generating and asking questions, and asks students to develop their own questions.</th>
<th>1. Teacher provides helpful suggestions and/or reflects on questions with students, other than simply providing the answers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teacher provides helpful suggestions and/or feedback on questions, other than simply providing the answers.</td>
<td>1. Teacher provides helpful suggestions and/or feedback on questions, other than simply providing the answers.</td>
<td>1. Teacher provides helpful suggestions and/or feedback on questions, other than simply providing the answers.</td>
<td>1. Teacher provides helpful suggestions and/or feedback on questions, other than simply providing the answers.</td>
</tr>
<tr>
<td>1. Teacher models the thought process for generating and asking questions, and asks students to develop their own questions.</td>
<td>1. Teacher models the thought process for generating and asking questions, and asks students to develop their own questions.</td>
<td>1. Teacher models the thought process for generating and asking questions, and asks students to develop their own questions.</td>
<td>1. Teacher provides helpful suggestions and/or feedback on questions, other than simply providing the answers.</td>
</tr>
<tr>
<td>1. Teacher models the thought process for generating and asking questions, and asks students to develop their own questions.</td>
<td>1. Teacher models the thought process for generating and asking questions, and asks students to develop their own questions.</td>
<td>1. Teacher models the thought process for generating and asking questions, and asks students to develop their own questions.</td>
<td>1. Teacher provides helpful suggestions and/or feedback on questions, other than simply providing the answers.</td>
</tr>
</tbody>
</table>

### Initial Thinking

<table>
<thead>
<tr>
<th>1. Teacher does not ask questions that enable students beyond their initial understanding.</th>
<th>1. Teacher asks questions that enable students beyond their initial understanding.</th>
<th>1. Teacher asks questions that enable students beyond their initial understanding.</th>
<th>1. Teacher asks questions that enable students beyond their initial understanding.</th>
<th>1. Teacher asks questions that enable students beyond their initial understanding.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teacher asks questions that enable students beyond their initial understanding.</td>
<td>1. Teacher asks questions that enable students beyond their initial understanding.</td>
<td>1. Teacher asks questions that enable students beyond their initial understanding.</td>
<td>1. Teacher asks questions that enable students beyond their initial understanding.</td>
<td>1. Teacher asks questions that enable students beyond their initial understanding.</td>
</tr>
</tbody>
</table>

### Types of Thinking

<table>
<thead>
<tr>
<th>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</th>
<th>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</th>
<th>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</th>
<th>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</th>
<th>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
</tr>
</tbody>
</table>

### Teach 6 – Check for Understanding

<table>
<thead>
<tr>
<th>Description</th>
<th>Level 1 – Significantly Below Expectations</th>
<th>Level 2 – Below Expectations</th>
<th>Level 3 – Meeting Expectations</th>
<th>Level 4 – Above Expectations</th>
<th>Level 5 – Significantly Above Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address Misunderstandings</td>
<td>1. Teacher attempts to check for understanding of content, but misunderstandings are not addressed.</td>
<td>1. Teacher attempts to check for understanding of content by addressing misunderstandings using the same approach/strategy.</td>
<td>1. Teacher checks for understanding of content by addressing misunderstandings with another approach/strategy.</td>
<td>1. Evidence fully supporting level 5 is present, as well as one of the following: 1. Teacher checks for understanding of content by allowing students to offer specific and relevant feedback to each other.</td>
<td>1. Evidence fully supporting level 5 is present, as well as one of the following: 1. Teacher checks for understanding of content by allowing students to offer specific and relevant feedback to each other.</td>
</tr>
<tr>
<td>Teacher Clues</td>
<td>1. Teacher invites students to share what they understand right away.</td>
<td>1. Teacher formally assesses students’ work without providing feedback in real-time (when needed).</td>
<td>1. Teacher formally assesses students’ work without providing feedback in real-time (when needed).</td>
<td>1. Teacher formally assesses students’ work without providing feedback in real-time (when needed).</td>
<td>1. Teacher formally assesses students’ work without providing feedback in real-time (when needed).</td>
</tr>
<tr>
<td>Format &amp; Assessment</td>
<td>1. Teacher uses a variety of methods to check for understanding.</td>
<td>1. Teacher uses a variety of methods to check for understanding.</td>
<td>1. Teacher uses a variety of methods to check for understanding.</td>
<td>1. Teacher uses a variety of methods to check for understanding.</td>
<td>1. Teacher uses a variety of methods to check for understanding.</td>
</tr>
<tr>
<td>Scoring Techniques</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
<td>1. Teacher’s failure to teach one type of thinking is inappropriate/redundant, low-level and/or inappropriate for the lesson.</td>
</tr>
<tr>
<td>Continuity</td>
<td>1. Flow of the lesson is not maintained; students who do not understand are constantly dropped.</td>
<td>1. Teacher is able to address student misunderstandings effectively, taking away from the flow of the lesson and losing the engagement of students who do understand.</td>
<td>1. Teacher is able to address student misunderstandings effectively, taking away from the flow of the lesson and losing the engagement of students who do understand.</td>
<td>1. Teacher is able to address student misunderstandings effectively, taking away from the flow of the lesson and losing the engagement of students who do understand.</td>
<td>1. Teacher is able to address student misunderstandings effectively, taking away from the flow of the lesson and losing the engagement of students who do understand.</td>
</tr>
<tr>
<td>Check for Understanding</td>
<td>1. Teacher utilizes only one method to check for understanding.</td>
<td>1. Teacher utilizes a variety of methods to check for understanding but fails to identify students who do not understand.</td>
<td>1. Teacher utilizes a variety of methods to check for understanding but fails to identify students who do not understand.</td>
<td>1. Teacher utilizes a variety of methods to check for understanding but fails to identify students who do not understand.</td>
<td>1. Teacher utilizes a variety of methods to check for understanding but fails to identify students who do not understand.</td>
</tr>
<tr>
<td>Description</td>
<td>Level 1 – Significantly Below Expectations</td>
<td>Level 2 – Below Expectations</td>
<td>Level 3 – Meeting Expectations</td>
<td>Level 4 – Above Expectations</td>
<td>Level 5 – Significantly Above Expectations</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>----------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Materials Prepared</td>
<td>1. Instructional materials are not prepared by the start of class.</td>
<td>1. Teacher attempts to maximize instructional time by having instructional materials prepared.</td>
<td>1. Teacher has instructional materials prepared by the start of the class.</td>
<td>1. Evidence fully supporting level 3 is present, as well as some of the following:</td>
<td>1. Evidence fully supporting level 3 is present, as well as all of the following:</td>
</tr>
<tr>
<td>Student Wait Time</td>
<td>2. Instructional time is not used effectively; leaving students idle for significant periods while waiting for the teacher.</td>
<td>2. Teacher does not maximize students’ wait time; they may be idle for short periods while waiting for the teacher.</td>
<td>2. Teacher minimizes students’ wait time.</td>
<td>2. Teacher provides opportunities for students who finish work early to engage in meaningful activities that extend self-guided learning.</td>
<td>2. Teacher provides opportunities for students who finish work early to engage in meaningful activities that extend and self-guided learning.</td>
</tr>
<tr>
<td>Timing</td>
<td>3. Teacher spends an inappropriate amount of time on one part of the lesson when students have mastered the objectives or demonstrated understanding.</td>
<td>3. Teacher spends an inappropriate amount of time on one part of the lesson when students have demonstrated their ability to move on to the next component of the lesson.</td>
<td>3. Teacher spends an appropriate amount of time on each component of the lesson.</td>
<td>3. Teacher structures the lesson so that student transitions do not interrupt the flow of the lesson.</td>
<td>3. Teacher structures the lesson so that student transitions do not interrupt the flow of the lesson.</td>
</tr>
<tr>
<td>Pacing</td>
<td>4. Teacher deviates the lesson at a notably slow pace that leaves students completely disengaged without anything meaningful to do.</td>
<td>4. Teacher executes the lesson at a pace that leaves students sometimes disengaged or without anything meaningful to do.</td>
<td>4. Teacher executes a coherently structured lesson that is appropriately paced, such that students are almost never disengaged or left without anything meaningful to do.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**The lesson’s pacing speed should be appropriate to the content covered, instructional strategies used, and lesson adjustments made based on a teacher’s checks for understanding.**