10-2-2013

Development and Use of Essential Learning Goals and Their Effect on Student Achievement in Grades Two through Five

Development and Use of Essential Learning Goals and their Effect on Student Reading Achievement in Grades Two through Five.

by

Scott Allen Andrews
M.Ed., Education Administration, University of Missouri—Saint Louis, 1998
M.Ed., Special Education, University of Missouri—Saint Louis, 1998

Submitted to the Graduate School of the

UNIVERSITY OF MISSOURI - ST. LOUIS
In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF EDUCATION

in

Education Administration

May 2011

Advisory Committee

Carole H. Murphy, Ed.D., Chairperson

Helene J. Sherman, Ed.D. Kathleen S. Brown, Ph.D. Dr. Cody Ding, Ph.D.

Abstract

This study looks at the use of Essential Learning Goals and their effect on student learning in grades two through five. Teachers in the treatment group participated in a yearlong professional development program. The treatment incorporated the concepts developed by Wiggins and McTighe (2011) in their research "Understanding by Design", Marzano's (2009) work on development of learning goals and objectives, and Hess's (2007) work on Learning Progressions.

The treatment provided training to teachers through a professional development program designed to enhance teachers' content knowledge to improve student achievement. Student achievement was measured using a district wide communication arts assessment tool. Communication arts scores were evaluated on fourteen different data points over a two year period. Scores were evaluated to determine if an effect occurred related to student achievement after teachers participated in the professional development treatment. Scores were collected using the E-valuate electronic assessment tool.

Results indicated that during the baseline year, student achievement scores improved in a similar manner. During the treatment year, the mean score for the control group increased by 2.27 points and the mean score for the treatment group increased by 12.57 points.

The difference in the growth of the scores between the control and experience groups was significant. An effect on student achievement scores occurred in the experiment group. The covariates of observation of goal use, teacher experience, and education beyond a bachelors' degree did not impact the degree of the effect occurring in the student reading achievement scores. It is the recommendation of the researcher additional research take place to confirm results and address limitations in this study.

Acknowledgments

I want to say thank you to my Doctoral Advisor Dr. Carol Murphy. She never gave up on me and has been the compass keeping me headed in the right direction over these oh so many years. Having to take a few years off to take care of family, she welcomed me back with a smile and her friendly hello. She helped me to get back on track and then kept up with encouragement and advice throughout the process. I cannot thank her enough.

I would also like to thank Dr. Cody Ding, Dr. Kathleen Brown, and Dr. Helene Sherman, the members of my committee. They were so gracious to sit on my committee and provide the guidance and encouragement they were so willing to give. Each and every one of them has provided me with ideas and things to consider along the way. Everything they asked me to consider helped to make my efforts more focused and improved how I was thinking about my topic. The all have made such a difference in my experience.

I cannot say enough about the support of my wife. For all of the times she had to be a single parent while I attended the night classes or was in the library working on research or writing my paper. She has been encouraging and never complained once during the many hours and days away from home. Her love and compassion have helped so much to keep me going and get this done.

To my kids who have put up with an absent father over the years of time spent in acquiring two masters' degrees and finally my doctoral degree. I apologize for all the times I missed an important event in your life or was not there to help with homework, play a game or tuck you in at night. I love all three of you.

Table of Contents

Chapter 1

Introduction

The United States Congress passed the No Child Left Behind Act (NCLB) which may have encouraged reform measures contributing to improvement in student achievement across the country. One of the most well-known provisions is the requirement for each state to develop and use an assessment process to measure growth in student achievement (U.S. Department of Education, 2009). These high stakes assessments are utilized as an accountability tool to determine which schools are failing and also as a summative assessment tool to provide a snapshot of student achievement. As a result, teachers across the United States are focused on providing data to support student performance. Whereas student achievement scores are improving to a degree, achievement scores across the nation are not keeping pace with the increasing accountability measures of NCLB (U.S. Department of Education, 2009).

In light of this emphasis on US student academic success, a possible contributing factor to the less than satisfactory student achievement is the national emphasis on high stakes testing. The use of assessment as an accountability and summative tool may come at the cost of using assessment as a formative tool and may also reduce the amount of time teachers spend developing their expertise in their content knowledge. Teachers' content knowledge has been noted as a crucial factor affecting student learning (Stiggins, 2002). A teacher's content knowledge and ability to determine learning goals are the first steps of the instructional process. It is necessary for teachers to have expert content knowledge to facilitate the determination of what students currently know, and then determine the next instructional steps (Stiggins, 2002).

Federal and state money continues to be spent in an effort to improve teachers'

instructional capacity. According to researchers, the nature of this professional development should be targeted to teachers learning specific skills directly related to their instruction (with the learning occurring over a sustained period of time) and allowing teachers time to practice said skills, receive feedback, and be provided additional support in their efforts to improve (National Staff Development Council, 2001). (See Appendix B on page 91for standards of High Quality Professional Development)

The process of becoming an expert in content knowledge to facilitate the development of lesson goals is difficult. The development of this content knowledge must be nurtured through a commitment to lifelong learning. For teachers to develop their content knowledge and their ability to determine learning goals, they must be supported through ongoing professional development (DuFour & Eaker, 1998).

For the purpose of this study, content knowledge will be referred to as *essential learning goals*. These essential learning goals encompass the enduring understanding, essential questions, goals, and objectives addressed in the constructivist approach of Wiggins and McTighe (2011). Also associated with the concept of content knowledge is the concept of learning progressions as outlined by Corcoran, Mosher, Rogat, and the Consortium for Policy Research in Education (2009). Learning Progressions are the pathways students take in the process of learning new content or skills. This includes each individual step a student must master to display a new skill or each bit of knowledge the student must know to display learning of a concept (Corcoran et al., 2009). In addition, the theories and work of Hess (2007) will be utilized to develop teacher knowledge of learning progressions.

The importance of determining what we want students to know has been addressed since the mid twentieth century (Thompson & Education Testing, 2009). After decades of dialogue

addressing the competencies necessary for teachers to effectively instruct students, including their ability to determine what students should know (essential learning goals), how teachers know what students have learned (assessment), and what impact information from assessment will have on further instruction (next instructional steps), teachers across the country still need to improve their ability to determine essential learning goals (Black & Wiliam, 2010; Wiggins & McTighe, 2008).

It is important to determine what impact the lack of development of essential learning goals has on student achievement. Although many teachers across the country are working hard in an attempt to meet the increasing achievement standards of NCLB, it is the author's position that teachers must also have a strong knowledge of the curriculum they are responsible to teach. Because the goal of instruction is to reach a measurable degree of learning or proficiency in a specific skill, the nature of learning or elements of a specific skill must be well understood by the teacher. Teachers profit from discussions determining which essential learning goals are relevant to the curriculum they teach. Important reasons include teachers need to determine a course of focused instruction to meet learning goals; assessment of students' learning progression; provide feedback to encourage students to think critically about their own learning; and finally, assessing student success in the learning of the essential learning goals of the course (Black & Wiliam, 2010; Hess, 2007; Wiggins & McTighe, 2011).

The development of essential learning goals is an instructional strategy that can be utilized across all curricular areas. It is as equally important for a teacher in art, music, or physical education to possess an expertise in the content knowledge of the curriculum they teach as it is for a teacher in communication arts.

Teachers' capacity to develop effective and accurate essential learning goals could impact

the use of several effective instructional strategies: teachers can provide students with the goals of learning prior to instruction; focus classroom instruction to stated learning goals; provide more focused feedback to the student; and improvement of teacher-created assessments to measure student learning of the essential learning goals (DuFour & Eaker, 1998; Stiggins, 2002). In addition, improving teachers' ability to determine essential learning goals may improve the quality of questions in that they require more than a yes or no answer and facilitate students thinking critically about their own learning progress (Black, Harrison, Lee, Marshal, & Wiliam, 2004).

If research can highlight that student achievement is increased by improving teachers' capacity to develop and understand the essential learning goals of the curriculum taught, perhaps education policy will change. The newly identified focus would be in supporting our teachers with the time, resources, and professional development necessary to successfully develop instructional capacities such as the development of essential learning goals.

Relevance of Study

The key elements of the curriculum organization tools *Understanding by Design* and *Learning Progressions* were used to develop the concept of essential learning goals for this study due their importance in the instructional process. Several instructional competencies emerged as important in the literature review to address the development and implementation of instruction. The need for a strong knowledge of content and the ability to develop essential learning goals are common themes related to each of the competencies discussed in the literature. These competencies include content knowledge, an ability to determine essential learning outcomes, knowledge of effective pedagogy, developing assessments, determining the appropriate assessment method, and administering, scoring, and interpreting assessment results. Additional

competencies include using assessment results to make instructional decisions, inform the feedback process, assign grades, and understand the ethical, legal, and appropriate use of the assessment (see Appendix A on page 90 list of competencies). Seven of the ten competencies were identified by three education groups in addressing the assessment competencies of the instruction process. The three education groups titled these competencies the Standards for Teacher Competence in Educational Assessment of Students (National Council on Measurement in Education, American Federation of Teachers, & National Education Association, 1990). Seven of the 10 instruction competencies are related to a teacher's content knowledge and ability to determine the essential learning goals: knowledge of effective pedagogy, developing assessments, determining the appropriate assessment method, and administering, scoring, and interpreting assessment results (DuFour & Eaker, 1998; Stiggins, 2002; Wiggins & McTighe, 2011). Additional competencies reliant on teacher knowledge and the ability to determine essential learning goals include using assessment results to make instruction decisions, inform the feedback process, and assign grades.

According to DuFour, Eaker, and DuFour (2005), teachers are not provided enough time to allow them to collaborate in teams to address the development of essential learning goals. Richard DuFour (2011) asserted that teachers in the United States spend more time in direct contact teaching students as compared with teachers in most other developed countries in the world. This direct contact results in less time collaborating on the development of instruction. DuFour (2011) also asserted that, within the professional learning communities' model, the first step in providing effective instruction is for teachers to collaborate and engage in discussion in order to determine essential learning goals. Black and Wiliam (2010) question the present practice of minimal planning time suggesting that it results in our teachers having to go it alone

with little to no support in completing the planning work they need to accomplish to implement their classroom activities. DuFour, DuFour, Eaker & Many (2006) also questioned why the work of identifying goal and objectives was not getting done.

Accepting assertions by education researchers (Black & Wiliam, 2010; DuFour, Eaker, & DuFour, 2005; Stiggins, 2002), the use of high stakes assessments is resulting in teachers focusing their efforts on improving student scores at the cost developing their own essential learning goals. A contradiction appears to exist in terms of what teachers should focus on to assure quality instruction versus what is actually taking place. Is this contradiction in the actual versus ideal practice of our teachers resulting in lower student achievement?

Research Questions

This study focused on the following question: If second through fifth grade teachers were supported with time and professional development to improve their competency in the development of essential learning goals in the area of reading, would their development result in improved student achievement?

In order to obtain data related to the stated problem and research question, the following hypotheses were developed:

Alternate Hypothesis

H_1 The development of reading essential learning goals in grades two through five will impact student achievement

Null Hypothesis

H_o The development of reading essential learning goals in grade two through five has no impact student achievement.

The Problem

Whereas educational researchers have studied the importance of teacher engagement in professional development related to increasing teacher capacity to determine the goals of instruction (DuFour & Eaker, 1998; Marzano, 2009) teachers across the country are reporting they are not engaged in determining what they want their students to know in a collaborative, organized manner (DuFour, 2011). Teachers attempting the difficult task of developing essential learning goals report that they do not feel comfortable in developing goals on their own and that they are receiving minimal support in these efforts through professional development with curriculum experts (DuFour & Eaker, 1998). Wiggins and McTighe (2011) asserted teachers should be provided with support in developing the skills necessary to effectively develop essential learning goals. DuFour, Eaker, and DuFour (2005) research concurs with Wiggins and McTighe (2011) the work of determining what students need to know should be completed with teachers working collaboratively and being given enough time and support to allow for rich and complete dialogue.

Issues arise regarding situations in which teachers need to develop essential learning goals and they may not have sufficient time and support to do so. Student achievement is not improving at the desired rate set forth in NCLB or as measured on international achievement assessments (Petrilli, Scull, & Fordham, 2011). Also, despite research indicating the need for teachers to possess the ability to determine essential learning goals, teachers are not provided professional development support or time to develop the essential learning goals necessary to facilitate the delivery of effective instruction (DuFour & Eaker, 1998). Is a lack of development of essential learning goals one of the factors responsible for the lack of increase in students' academic achievement?

The purpose of this study was to determine if providing teachers in grades two through five with professional development support and collaborative time to develop essential learning goals will impact student achievement. Research indicates that teachers are experiencing difficulty in developing essential learning goals for a variety of reasons (Wiggins & McTighe, 2008).

In this study, teachers in grades two through five were provided with additional collaboration time and professional development support to focus on the development of essential learning goals in the area of reading. The professional develop process targeted specific skills necessary in the development of essential learning goals over a sustained period.

The study also investigated possible effects of three teacher descriptor related factors on student achievement: years of teaching experience; years of education past a bachelor's degree; and observations to determine the use of essential learning goals in the classroom.

Delimitations

Delimitations of the study include the following:

1. The study period was two years year.
2. The study sample is a medium-sized school district in the Midwest.
3. The study sample includes teachers at two small suburban elementary schools.
4. The study examines if essential learning goals, questions, goals, and learning progressions can be created given support, and if any impact on student achievement is realized.

Limitations

Limitations of the study include the following:

1. Participation in the study among a limited number of teachers resulting in a small data set.

2. Teacher participants' motivation may be influenced by their relationship with the researcher.
3. The teacher participants' may put forth more effort than would be normal due to the subordinate relationship with the researcher. Motivation factor irregularities may cause the results to be atypical.
4. The small size of the data set will make generalization to the wider education community difficult.
5. The limited nature of the student socio-demographic characteristics producing the data will make generalization to the wider education community difficult.

Assumptions

Assumptions in the study include the following:

1. Teacher participants in the study will comply with request to post essential learning goals due to the researcher's supervisory relationship with the teacher participants.
2. When being observed by the researcher, the teacher participants will refer to essential learning goals in the lesson due to the researcher's supervisory relationship with the teacher participants.
3. Teacher participants will participant with best effort in the professional development treatment activities.
4. Teacher participants will engage in all professional development activities.
5. Student assessments scores will be representative of best possible student effort.
6. Assessment of students will be conducted in a quiet environment; students will not experience interruptions during the assessment, and all students will be provided enough time to complete the assessment in its entirety.

Definition of Terms

For the purposes of this study, several terms are defined below.

Formative Assessment - this phrase refers to actions taken by a teacher to acquire information concerning a student's knowledge of desired goals or objectives. The information gained by the assessments is utilized to inform instruction, to determine next steps for the student, and to provide information to be used as feedback to modify teaching and learning activities. In simple terms, assessment becomes formative assessment when knowledge gained about a student is used to adapt the teaching to meet the student's needs (Black & Wiliam, 2010).

Enduring Understandings - the big ideas or important overarching concepts students will learn. Enduring understandings refers to the ideas or concepts the student remembers when the small details of the content have been forgotten (Wiggins & McTighe, 2011).

Essential Questions - Questions that lead students to ask themselves questions in order to make meaning of the content they are taught. These questions elicit an internal process causing the student to think critically about their own learning (Wiggins and McTighe, 2011).

Learning Goals/Objectives - statements of the knowledge or skills students should possess and be able to do at the end of a unit of study or course. The goals and objectives should be measurable, specific, and easily defined (Marzano, 2009).

Learning Progressions - pathways students take in the process of learning new content or skills (Corcoran, Mosher, Rogat, & Consortium for Policy Research in Education, 2009).

Essential Learning Goals - the term used, in this paper, to include the four curricular content constructs of enduring understandings, essential questions, goals, and objectives, and learning progressions to define the components making up a teacher's subject content knowledge.

Organization of the Study

The study is organized into five chapters, a bibliography, and appendices in the following order: Chapter 2 presents a review of the literature highlighting a history of assessment in the education field, the impact of legislation on the field assessment, assessment competencies recommended in the literature for use in the classroom, and professional development issues surrounding improvement in the instructional practice of educators. Chapter 3 describes the research design of the study along with the methodology of the study, including the instruments utilized to gather data for the study, the procedures followed in the study; and the process for determining the study sample. Chapter 4 presents a discussion of the data analysis, and Chapter 5 contains summary, conclusions, and recommendations made based on the study results.

Chapter II

Literature Review

Introduction

Results of high stakes assessments and international tests of student achievement indicate students in the United States are not performing well relative to their peers in other developed nations. Results indicate students in the United States are scoring consistently lower in the areas of mathematics and science than students in other developed nations as measured by the Third International Math and Science Tests (Mullis, Martin, Gonzalez, & Chrostowski, 2004).

The passage of the No Child Left Behind Act (NCLB) was an attempt to address k-12 public education students' academic achievement. NCLB established learning standards and an assessment system to measure mastery (Wylie, Lyon, Goe, & Educational Testing, 2009). Other efforts to reform assessment instructional practice and teacher quality have taken place to improve student achievement in this country.

One of the areas identified as crucial in assuring teacher quality is the degree of content knowledge a teacher possess related to the curricula they are responsible to teach. According to the Department of Education (2008), states must "collect data and monitor performance to ensure that all states meet the goal of having all core academic classes taught by highly qualified teachers in the school year 2006–07 and beyond" (p. 5). Each of the 50 states has developed guidelines to work towards meeting the NCLB mandate of hiring a High Quality Teacher for every classroom. The level of a teacher's content knowledge in a key component of that mandate (Department of Education, 2009).

In 2007, the state of Missouri Department of Elementary and Secondary Education (DESE) began developing the Missouri Model Teacher and Leader Standards: A Resource for State

Dialogue to address teacher quality. The standards are aligned to the Interstate New Teacher Assessment and Support Council Standards. These standards draw upon research utilized in the creation of the common core standards in the areas of English language arts and mathematics (DESE, 2012). The standards are also aligned to several other professional education associations' standards such as the National Council of Accreditation for Teacher Education (NCATE), the National Staff Development Council (NSDC), and the National Board for Professional Teaching Standards (NBPTS) (DESE, 2012). This work began with the Advisory Council of Certification for Educators and lasted from 2007 until 2009. Eventually 18 different organizations and associations, 32 school districts, and 25 Missouri institutions of higher education would partner in contributing to the creation of the standards. These standards were created to align with the framework outlined in Missouri Senate Bill 291. The standards were not created to be utilized as a checklist for school district to follow, but rather as a tool to facilitate dialogue within school districts in the creation of their own teacher quality standards (DESE, 2012). Senate Bill 291 provided guidance on appropriate standards divided into six general concept statements, nine teacher quality standards, and 36 quality indicators. The standards were approved by the Missouri State Board of Education in 2011 (DESE, 2011).

Teachers's content knowledge is the first teacher competency addressed within the nine quality standards developed in the DESE document. Standard one addresses, "content knowledge and perspectives aligned with appropriate instruction" (DESE, 2013b, p.1). Tied to the standard of content knowledge is "the teachers understands the central concepts, structures, and tools on inquiry of the discipline(s) and creates learning experiences that make these aspects of subject matter meaningful and engaging for student" (DESE, 2013b, p.1). These two concepts connect directly content knowledge and elements of instruction.

Rubrics were created outlining the skills a teacher should possess within each quality indicator. Each of the 36 quality indicators is comprised of five levels of proficiency. These five indicator proficiency levels include Candidate, New Teacher, Developing Teacher, Proficient Teacher, and Distinguished Teacher (DESE, 2013). The distinguished indicator for standard 1 content knowledge states "the distinguished teacher also…has mastery of taught subjects and continually infuses new research-based content knowledge into instruction" (DESE, 2013a, p.8). Each proficiency level is supported by "professional frames" (DESE, 2013a, p. 8). The three frames are evidence of commitment, evidence of practice, and evidence of impact. Going from new to distinguished in the rubric for content knowledge, commitment is as follows:

> Is well prepared to guide students in a deeper understanding of content; stays current on new content and incorporates it into lessons; use of supplemental primary sources that are aligned to local standards; and continually expands knowledge base on content and infuses into content. (DESE, 2013b, p. 8)

In an ETS study, *How Teaching Matters*, it was concluded that student grade level achievement increased by 40 percent in math and science if a teacher majored or minored in the content area for with the teacher provided instruction (Wenglinsky, 2000). Wenglinsky (2000) did find two other teacher competencies significant to improved student academic success, instructional practices/methods and professional development. Wiggins' and McTighe (2008) arguments concur with Wenglinsky's (2000) concerning the importance of classroom instructional practices, but suggest that strong content knowledge is key to a teachers ability to utilize the most effective instructional strategies to meet the goals of a lesson.

With the importance of a teacher's content knowledge clearly acknowledged among many professional education associations or organizations, it is important to understand the impact of a

teacher's solid foundation of content knowledge. Also, it is critical to analyze possible causes for a lack of focus on developing our teachers' content knowledge when it appears logical to do so. Assessment as an education tool is misused (Stiggins, 2002), and at times the competencies necessary for a teacher to effectively implement formative assessment are lost to the process itself (Wiggins & McTighe, 2008).

This paper focused on the following research streams: a look at the nation's perseveration with assessment as an educational tool, the instructional capacities within the formative assessment process identified by professional educations organizations as keys to effective and efficient classroom instruction, and a discussion of the lack of review in the literature on a systematic professional development plan to address improving the instructional capacities identified in the literature as necessary to implement formative assessment. The process of identifying what students should know, assessing that knowledge, and developing an appropriate plan of instruction were argued to be key instructional components by DuFour, DuFour, Eaker & Many **(**2006).

This paper will also address research streams on increased professional development; outline what research identifies as quality professional development, and review how content knowledge impacts teachers' ability to carry out the formative assessment process.

Chronological History of Assessment

Assessment in the early 1940's. Few assessment tools were utilized seventy years ago to inform instruction or to hold public schools or school systems accountable for student achievement. Rather, students were assessed for their learning aptitude (Department of Education, 2009). Tests were used to group students based on their assumed ability to learn, with teachers labeling students from slow to fast. Teachers grouped/tracked students according to

their perceived ability to learn, and some teachers even questioned if the slower/lower learners could ever learn the expected curriculum (Lewis, 2006). This model of grouping and tracking students tended to have a negative effect on poor and minority students, but grew in practice due to the pressure to educate a large influx of immigrants. The use of IQ testing was indicative of the prevailing beliefs about the varying abilities of the diverse population of learners in the United States (Lewis, 2006).

During this time, aptitude testing was viewed as a scientific tool. The process of tracking learners by IQ was related to teacher expectations. Some students were expected to learn very little, some were expected to learn an adequate amount, and others were expected to learn a great deal of the content. As it became clear that using assessment to place students in different learning tracks was harmful rather than beneficial in promoting the learning of all students, educators began looking for better assessment methods to promote learning (Lewis, 2006).

In the mid 1900s, assessment began to be discussed as a tool to not only measure learning, but to serve as a guide to promote student learning. In reference to outcome based learning, Guskey (1994) stated, "its guiding principles were elegantly set forth in the 1940s by Ralph W. Tyler in his classic book, *Basic Principals of Curriculum and Instruction*" (para. 2). According to Guskey (1994, para. 2), Tyler, an education researcher in the 1940s, addressed several questions related to assessment and pedagogy.

1. What educational purposes should the schools seek to attain?
2. What educational experiences can be provided that are likely to accomplish these purposes?
3. How can these educational experiences be effectively organized?
4. How can we determine whether these purposes are being attained?

Guskey (1994) stated that, "Tyler's approaches were referred to as objective-based education and were popularized because the 'back to basics' movement then dominated American education" (para. 5). Tyler's objective-based work in the 1940s was similar to the mastery movement of the 1960s and 1970s as well as the early stages of formative assessment as we know it today. However, the objective-based approach did not give much attention to the instructional methodology components in Tyler's second and third questions. In addition, Tyler's work did not focus on the use of the information gained from assessment to inform instruction (Guskey, 1994). The literature indicates that Tyler's objective-based methodology focused primarily on curriculum and assessment, or his first and fourth fundamental questions.

Guskey (1994) further argued that the focus on Tyler's first and fourth questions in his objective-based approach resulted in an instructional practice sounding very familiar to the mastery learning approach. Guskey (1994) states, "Under objective-based approaches, complex learning tasks were broken into smaller, more basic skills which then were arranged in an appropriate sequence for students to learn" (para.5). This approach reduced the learning and instruction process to a series of smaller elements for which little discussion occurred regarding the best methodology to facilitate student learning. Educators grew tired of an instructional methodology that was neither effective nor efficient in promoting student learning, and began to seek a more effective method of instruction. These positions lead to change in order to keep pace with the current methodology (Guskey, 1994).

Assessment in the 1960's. Scriven is considered to be one of the first to discuss the use of assessment in a formative nature (Black & Wiliam, 1998). According to Black and Wiliam (1998), Scriven determined that formative assessment methods could be applied to determine the worthiness of professional curricular materials. His discussion surrounded the assessment of

learning outcomes as a process for determining whether an education authority should allocate taxpayer funds to one product or another. He identified an evaluation process requiring ongoing assessment to determine the effectiveness of the curricular materials.

Assessment in the 1970's. In the early 1970s, a new movement requiring students to pass exit exams or standardized tests to graduate high school began to find favor within the education community. This new requirement was in contrast to earlier expectations that students simply maintain good grades and stay out of the office for disciplinary issues (Warren & Grodsky, 2009). During the early 1970s, some argued that assessment should focus on a set of basic skills while others asserted that it should focus on an increased set of rigorous academic skills necessary for entrance into college or directly into the work force (Lewis, 2006). Both assertions require that exit exams be given to students in order to measure mastery of minimum competencies. Warren and Grodsky (2009) argued the following:

> Proponents of exit exam policies say too many students simply get credit for 'seat time,' graduating without basic literacy and numeracy skills. With the decline in manufacturing and growth of the information economy, architects of exit exam policies have sought to bolster the value of the diploma. Supporters say these policies have increased pressure on students, parents, teachers, and school systems to boost academic achievement and to better prepare young people for college and the global economy. (p. 646)

These arguments supported the use of assessment to pressure school systems to improve academic performance but did little to address the use of assessment to improve instruction.

According to Warren and Grodsky (2009), critics of exit exam policies argued that

requirements actually harmed students failing the exams and did not benefit the students passing the exams. They further claim that exit exams are not nearly rigorous enough for the purpose of truly raising the standards of learning for students passing the exam, and only slightly reduce the number of students graduating. Warren and Grodsky (2009) also alleged that exit exams are rooted in the political rhetoric of accountability but do nothing to truly make a difference in student learning, and should be dropped if changes are not made to their current form. In a report presented at a National Center for Research on Evaluation, Standards, and Student Testing (CREEST) conference, Lewis's (2006) arguments were very similar to those of Warren and Grodsky (2009), alleging that exit exams do very little to raise standards and only work to exclude a small group from completing high school. Even with the arguments made concerning the effectiveness of exit exams, Warren and Grodsky (2009) contend that nearly two out of every three high school students today take some form of exit exam as a requirement to graduate.

While the practice of using exit exams began in the 1970s and still exists today, other uses and forms of assessment were developing by educators across the country during that same period. Rather than waiting to assess students' learning at the end of their educational career, some educators began utilizing assessments tools in the mastery of learning process. The hope was this process of learning would be effective and efficient, breaking the curriculum into tiny pieces and giving frequent assessments to students to evaluate their mastery of content. An early example of use of mastery learning was the Winnetka Plan utilized in Winnetka, Illinois. The Winnetka plan included the use of mimeographed worksheets with a narrow focus on curriculum skills. The students could take individual tests and then a test was given by the teacher when the student felt ready (Lewis, 2006). Lewis (2006) argued that breaking the curriculum down into tiny bits of learning and assessing students' mastery of each tiny bit was ineffective in promoting

student learning. Dividing broad goals and objectives into small, isolated sections, assessing learning of those separate pieces, and assuming students would then be able to reassemble those bits of knowledge to create a larger understanding proved to be incorrect at best. In fact, research shows that the process could actually hurt student learning. In addition Mastery Learning was an ineffective and inefficient form of assessment to promote student learning,

The Mastery Learning approach discussed by Lewis (2006) and the objective-based approach discussed by Guskey (1994) appear to be similar. Both assessment methods fell out of favor with educators due to teacher and student boredom and the lack of sustained improvement in student learning. Regardless of the terms utilized to define the process of determining goals and developing assessment, educators appeared to develop a negative association with the process of dividing content into smaller elements and assessing students on these smaller elements. According to Guskey (1994), educators developed negative connotations with the terminology such as competencies, objectives and goals to describe these small units of learning and continued to look for a term to define the expected content. The term "outcomes" found favor and was not associated with the negative connotation of previous terms.

In the 1970s the term formative and summative would no longer only be associated with assessment. The two terms would come to have separate meanings. The important change in meaning would be associated with the function of the assessment tool rather than assessment tools checking for mastery of content. One impact of this change would be teachers needing to change their assessment practices to keep their summative practices from impacting their new formative assessment tool (Black & Wiliam, 2003).

Assessment in the 1980's. In this period of time, assessment began to move forward as a tool promoting student learning consistent with the principles of formative

assessment. The fundamentals of formative assessment began to take shape, including: the determination of a clear definition or reference point relating to a set of goals; the instruction needed to convey the goals; the observations necessary to assess the degree of success in learning the goals; and finally, providing feedback to the student to influence the learner's future learning (Sadler, 1989). Adding the feedback step in the formative assessment process was a significant piece in facilitating student learning as an outcome of the assessment process itself. According to Sadler (1989), providing appropriate feedback to the student within the instruction process was an important factor in developing formative assessment as a tool for informing the teacher's instructional practice: Sadler (1989) stated:

> Feedback is a key element in formative assessment, and is usually defined in terms of information about how successfully something has been or is being done. Few physical, intellectual or social skills can be acquired satisfactorily simply through being told about them. Most require practice in a supportive environment which incorporates feedback loops. (p. 120)

Providing feedback to inform students of their level of understanding or ability related to a set standard is the essence of what makes formative assessment different from summative assessment (Sadler, 1989). It makes the assessment process an active one instead of a passive one, informing students on their degree of mastery as well as the next steps in the process to improved understanding or performance. Sadler (1989), asserted:

> "…the learner has to (a) possess a concept of the standard (or goal, or reference level) being aimed for, (b) compare the actual (or current) level of performance with the standard, and (c) engage in appropriate action which leads to closure of

the gap. (p. 121)

An additional change was including information gained in the assessment process to allow teachers to monitor their own effectiveness in terms of students' learning. Data affected instructional practice and engaged the students in their own learning, as well as thinking critically about content delivered. The ability of the teacher to carry out this important instructional process is reliant upon the degree of understanding or knowledge of curriculum content.

Assessment in the 1990's to current day. In the 1990s, the emphasis was on educators gaining information from assessments to provide feedback on their own effectiveness as well as the students' learning of identified goals and objectives. Black and Wiliam (1998) noted that formative assessment had two key distinguishing characteristics. The action to be taken by both the instructor and students as a result of the information learned; the assessment must result in teachers modifying their work due to the information learned from the assessment. A number of researchers have continued to redefine formative assessment, which Pryor and Crossouard (2008) consider to be an act of determining, teaching and assessing learning goals. The actions that need to be taken by teachers and students must be based upon the results of an assessment in order to further a student's learning and motivate and enhance student performance.

The basic guiding questions presented by Tyler nearly 70 years ago are very similar to the guiding questions addressed today within the Professional Learning Communities and Response to Intervention (RTI) movements. The guiding principles or questions addressed in both movements are directly related to the work cited by education practitioners (Eaker, DuFour, & DuFour, 2002) and education researchers (Black & Wiliam, 1998; Sadler, 1998; Pryor & Croussouard, 2008). In examining the fundamental questions asked today, it appears that it has taken nearly 70 years for the education community to clearly define an effective and efficient use

of assessment along with the instructional competencies to implement it effectively. Examining the three questions addressed in the PLC and RTI movements addresses the fundamental principles of formative assessment. In addressing Eaker, DuFour and DuFour's (2002) first question, 'what exactly do we expect students to learn?" (p.19); the process of determining content or curriculum goals and objectives to be covered is addressed. Curriculum, content, objectives, and goals have all been topics of discussion for the past 40 years, and this discussion continues among researchers today (Black & Wiliam, 2010; Guskey, 1994; Marzano, 2009; Sadler, 1998; Wiggins & McTighe, 2011). The second question, "How will we know what students are learning?" (Eaker et al., 2002, p. 19) addresses the assessment of what has been taught to the student. The third question sounds so simple "How can we assist and support students in their learning?" (Eaker et al., 2002, p. 19), but it has taken the education community years to define this issue in terms of finding an efficient and effective instructional process. In reviewing the literature (Black & Wiliam, 2010; Lewis, 2006; Pryor & Croussouard, 2008; Sadler, 1998), it appears teachers and administrators have addressed this third question only in the last two decades. The issues addressed include reviewing assessment data to determine what students have learned and still need to learn; determining necessary feedback to move students forward in their understanding and skills; and informing teachers of instructional changes needing to be made.

Today, the discourse in the education community closely links assessment and instruction methodology. In many of the workshops and professional development conferences available today, the topic of assessment and its link to the determination of instructional practices is addressed (Stiggins, 2002).

Academic change: Why so slow? With a focus on assessment to inform instruction in the education community and the availability of professional development available, why has it taken so long for wide spread implementation in classrooms across the country? One possible answer stems from the very nature of how change in the instructional process occurs within the education community itself. Historically, each popular movement in education has been succeeded by a new movement that is a refinement of its predecessor. In an effort to describe the professional development work supporting competency based learning, McCowan (1998) argues that "the origins of any educational movement are difficult to describe because theoretical concepts seldom have a direct, straight-line influence on related theories. Instead, they overlap; draw from each other, and change — sometimes in reasonably clear patterns, but often in erratic, unpredictable ways" (p. 6). Stiggins (2002) asserted that one additional cause for the erratic changes in the focus on instructional and assessment methods could be the influence of government actions and concerns of the business community in the United States. Both issues cause confusion and anxiety for the teacher in the classroom.

Instructional Capacities Facilitating Formative Assessment as an Instructional Tool

With over 70 years of theoretical study in the academic community, laws promoted by Congress and the Office of the President of the United States, and continuous reform initiatives by the education community itself, a question can be raised. Why have we, as a country, not experienced greater increases in student achievement through the utilization of assessment? All of the reform movements and new visions on the utilization of assessment for learning require teachers to make major changes to their instructional practices, but insufficient support is available to facilitate these necessary changes (Borko & Putnam, 1996).

Researchers and education practitioners have also defined the competencies associated with

the effective utilization of formative assessment and the instructional practices necessary to incorporate formative assessment (Bell and Cowie, 2001; Bennett & Cunningham, 2009; Black & Wiliam, 1998; Borko, 1997; DuFour & Eaker, 1998; Lewis, 2006; Popham, 2006; Sadler, 1989; Stiggins, 2002). In addition to articles on the steps necessary to effectively implement formative assessment, professional education organizations and associations have asserted a list of competencies necessary for the implementation of formative assessment (The National Council on Measurement in Education, American Federation of Teachers, & National Education Association, 1990). The list of competencies identified to carry out effective utilization of formative assessment and the instructional practices associated with the instructional tool include ideas related to content background and appropriate assessment connections. Specific competencies are listed in Appendix A on page 90.

The following sections discuss competencies recommended in the literature to carry out formative assessment and related to the teacher's knowledge of curriculum content. These competencies have been addressed in articles by a variety of researchers, educators, and professional education organizations and associations.

Knowledge of curriculum subject matter. Possessing a strong knowledge of curriculum content is a very important competency related to instructional practices associated with formative assessment. Within the formative assessment process, an educator's solid knowledge of the curriculum content is necessary in the first step of formative assessment: setting goals to determine what you want your students to know (Eaker, DuFour & DuFour, 2002). Teachers without substantial knowledge of the content they teach will not be successful in the first step of determining what they want their students to know.

The importance of content knowledge is also highlighted in an article by Rothman (2000)

on the mathematics content knowledge of elementary teachers in the United States versus elementary teachers in Asian nations as measured by performance on international achievement tests. Rothman (2000) discusses the inability of college students at the University of Wisconsin - Madison to solve mathematics problems that all elementary education teaching program students in Asian nations are expected to solve in college entrance exams. Based upon conversations with professors in education programs, Rothman (2000) raised concern students in teacher education programs in the United States do not possess the necessary content knowledge in the area of mathematics, facilitating an inability to be as effective and efficient in instruction leading to student learning.

In response to students lacking the necessary prerequisite skills in core academic areas, scholars are recommending changes to state certification testing programs (Rothman, 2000). Ginsburg, Leinwand, Anstrom & Pollock (2005) recommend that states utilize the PRAXIS II Multiple Subjects Assessment for Teachers, as it appears, to present an increase in the degree of challenge in content versus the typical PRAXIS II, which is currently required by states across the country. These concerns raised by education program professors and their call for an increase in testing for improved content knowledge among students exiting our nation's college education programs highlight the importance of the content knowledge of teachers. Borek (2008) asserted the importance of a teacher's grasp of content knowledge and subject matter in citing the Nation at Risk report. Finally, Lewis (2006) asserted the importance of linking a teacher's content knowledge to pedagogy and the need for ongoing study in the importance of our teachers' content knowledge.

Knowledge of the student. In their discussion on formative assessment, DuFour, DuFour, Eaker & Many (2006) asserted there are no easy shortcuts for implementing their

Professional Learning Community (PLC) process embracing formative assessment as the cornerstone of a PLC. Teachers in a PLC will address what they want their students to know—content knowledge—and second, determine what the student knows. Thompson et al. (2009) assert the necessity of the teacher to be able to assess the student's current degree of knowledge of learning goals to facilitate a response to that state in a manner promoting effective and efficient learning and achievement.

Ketterlin-Gelle (2005) stresses the importance of learner knowledge in assisting the teacher in two of the necessary steps of formative assessment: understanding what identified content goals students know and the actual development of the assessments. DuFour, Eaker, & DuFour (2005) argued for the importance of utilizing formative assessment over summative. Utilizing formative assessment shifted the focus from measuring performance to assessing student knowledge against a set of essential learning goals.

Without an extensive knowledge of the student, a teacher is unable to provide the necessary feedback that will allow students to think critically about their own learning. Stiggins, Arter, Chappuis, J. & Chappuis, S. (2005) asserted teachers utilize knowledge of their students, gained from assessment to design lessons that cause students to think critically about their own learning. They proposed that students needed to be engaged in their own learning through assessing themselves in three questions: (1) "Where am I going?" (2) "Where am I now?" and (3) "How can I close the gap?" (p.42). Stiggins et al., (2005) stated that to get students to think critically about their own work in this manner, the teacher must make sure their students know "the learning target (s), objective (s), or goal (s) in advance of teaching the lesson, giving the assignment, or doing the activity" (p.42). To implement these instructional strategies outlined by Stiggins et al., (2005) knowledge of content and understanding of concepts is a key component in

the competencies a teacher must possess to effectively and efficiently utilize assessment to improve student learning.

Determining essential learning goals Changes in a teacher's capacity to determine essential learning goals for the classroom could promote increased use of several effective instructional strategies: Teachers providing students with the goals of their learning prior to instruction beginning; teachers improved focus of classroom instruction to stated learning goals; an improvement in the focus of feedback to the student; and finally improvement in the development of teacher developed assessments to measure each student's learning of the essential learning goals (Stiggins, 2002). Wiggins and McTighe (2011) asserted the development of essential learning goals as the key to instruction success in the classroom. A teacher's ability to determine all of the steps of the instructional process are reliant upon a teacher's ability to determine the essential learning goals of the curriculum they are responsible to teach. Marzano (2009) considered development of solid learning goals and objectives as the key to student learning. If the teacher was not able to identify key learning goals and objectives, student learning would suffer.

Knowledge of effective pedagogy. In addition to hiring teachers who have a high level of content knowledge, are able to discern what their students should know, we must also find teachers with a strong understanding of pedagogy. The National Commission on Teaching and America's Future (1996), a 27 member commission including elected governmental officials, collegiate education professionals, K-12 professional education practitioners, and heads of major American corporations, conducted a 2-year study and developed recommendations for improving student achievement across the United States based upon three agreed upon premises. The first, and possibly the most important of the commission's premises, is "what teachers know and can

do is the most important influence on what students learn" (p. VI). Based upon this premise as well as the commission's third premise, "school reform cannot succeed unless it focuses on creating the conditions in which teachers can teach and teach well," (p. VI) the commission recommends that teacher education programs focus on content and pedagogy in pre-service programs, include methods and content in state teacher testing for certification, and provide teachers with continuous professional development in the craft of teaching. The commission also recommends standards for teaching should be a key element in the nation's efforts to reform education in the United States. Their report states: "evidence already exists that where school faculties are working together to translate standards into courses of study, learning tasks, and assessments, they are becoming more expert and more collective in their practice, and students are learning more" (p. 66).

Bloom (1968a) wrote, "Most students (perhaps over 90 percent) can master what we have to teach them, and it is the task of instruction to find the means which will enable our students to master the subject under consideration" (p. 1). Bloom (1968b) also spoke to the importance of assessment related to teacher competency in pedagogy as a key factor in the ability to develop lessons meeting the needs of each student. Shulman (1987) stressed the importance of teachers possessing strong content and pedagogical knowledge in the subject they teach, stressing that teachers need to match the instructional methodology to the content they are presenting in order to make the content easily comprehensible by the students.

According to Shulman (1987), there are three areas of expertise in instruction. The first area of expertise is content knowledge, the second is pedagogical knowledge, and the third is a combination of the first two areas.

Shulman (1987) also discussed the importance of the teacher possessing a strong

knowledge of her students' existing subject matter knowledge and general background knowledge, and recommends a five step process she refers to as "transformation" in the combining of content knowledge, pedagogical knowledge, and student knowledge.

Knowledge in Development of Appropriate Assessment Methods. On occasion, some parents question the purpose of a question on a test, but most assume that if the question is on the test it must be important. These weekly or end of unit tests can be useful formative assessment tools, but there needs to be a considerable amount of work completed by the teacher prior to and after the assessment is utilized in the classroom to assure the assessment is a functional tool informing instruction (DuFour et al., 2005). Rudner and Schafer (2002) stated that education researchers and practitioners were stressing the importance of assessment principles, techniques, and procedures that teachers and administrators should know to properly develop and utilize appropriate formative assessment methods. Rudner and Schafer (2002) also made recommendations on the necessary knowledge and understanding necessary for the development of effective and efficient assessment tools based on established professional assessment standards such as "Standards for Teacher Competence in Educational Assessment of Students (1990),… the Code of Professional Responsibilities in Educational Measurement (1995), the Code of Fair Testing Practices (1988), and the new edition of *Standards for Educational and Psychological Testing* (1999)" (p. 6) (See Appendix C on p. 94 for recommendations from Rudner and Schafer).

Rahn and Stecher, (1997) considered similar issues in their discussion of the knowledge necessary for developing appropriate assessment methods. According to Rahn et al. (1997), quality assessments must be reliable, valid, and fair. They also argue assessment tools need to be readily accessible to the teacher, and the teacher must be able to discern which assessment tool best facilitates integration with instruction. Finally, teachers must understand the information

gained from the results of the assessment tools and find the results credible in order to promote the use of the results to impact instruction. Rahn et al. (1997) argued that effectively developing or choosing an assessment method is critical in optimizing the results of an assessment model. The use of appropriate assessment methods to efficiently and effectively impact instruction is important, as is possessing the knowledge and skills necessary to develop appropriate assessment tools necessary to guide instruction (Stiggins, 2002). It is also necessary to provide systemic professional development to build capacity in teachers' ability to develop or choose appropriate assessment methods.

Assessment Instructional Strategies. Summative and formative assessments are those most often utilized. Summative assessment can be used to determine what the students are learning, and formative assessment can guide or inform instruction to facilitate learning. Summative assessment can be described as a snapshot in time as it reveals what the students know at a given point in time. Formative assessment can be described as a process by which the teacher utilizes the information gained from the assessment to further students' understanding of predetermined goals, objectives, or outcomes. Both forms of assessment serve an important function in the education process of students, but it is the proper utilization of formative assessment that effectively and efficiently impact students' achievement.

Black, Harrison, Lee, and Marshal (2004) stated effective utilization of formative assessment should change the instructional strategies incorporated in teachers' lesson planning. One of the most significant changes is the learning process becomes collaborative in nature. The teacher must accept the student as an active partner instead of simply being a recipient of knowledge. According to Black et al., (2004), to achieve this change instruction should provide the teacher with opportunities to listen to the students thinking; give feedback causing students to

think critically about and improve their own learning; utilize strategies that promote a collaborative learning environment; causes students to listen to each other critically, and foster their students willingness to express their understanding of the content being taught.

Heritage, Kim, Vendlinski, and Herman (2008) described the formative process as one in which teachers determined what students knew through various assessment methods, including quizzes, observations, formal written tests, and discussions with the student. These assessments can take place at predetermined intervals, intermittent intervals, or as an ongoing process during a teacher's lessons. Heritage et al. (2008) suggested teachers should examine or look for evidence of students' current understanding or misunderstanding of predetermined goals or objectives: "they need to infer the gap between the students' current learning and desired instructional goals, identifying students' emerging understanding or skills so that they can build on these by modifying instruction to facilitate growth" (p. 1).

As suggested by Black et al. (2004) and Sadler (1989), teachers should possess the ability to acquire knowledge of what their students know and understand, and accordingly adapt instruction to meet the needs of those students. Eaker, DuFour & DuFour (2002) stated the proper utilization of assessment includes the following: teachers collaborating in the development of quality assessments based upon predetermined goals and objectives; the setting of clear and accurate expectations aligned to learning goals for the students; and a systematic approach to the development of future instruction based upon information learned about the students from the gathered data.

Stiggins (2002) argued that to facilitate effective and efficient use of assessment as a learning tool, policy must change to include items such as the following: "match every dollar invested in instruments and procedures intended for assessment of learning at national, state, and

local levels with another dollar devoted to the development of assessment for learning" (para. 54). Stiggins (2002) further calls for the "launch a comprehensive, long-term professional development program at the national, state, and local levels to foster literacy in classroom assessment for teachers, allocating sufficient resources to provide them with the opportunity to learn and grow professionally" (para. 54). Stiggins (2002) also called for states to "change teacher and administrator licensing standards in every state and in all national certification contexts to reflect an expectation of competence in assessment both of and for learning; and require all teacher and administrator preparation programs to ensure that graduates are assessment literate -- in terms both of promoting and of documenting student learning" (para. 54).

Administering, Scoring, and Interpreting External/Internal Formative Assessments

Assessments must be administered in an environment that is fair and consistent to give students the opportunity to complete the assessment task in a risk-free environment. It is the teacher's responsibility to be aware of any and all accommodations needed to facilitate an assessment environment conducive to the student's successful attempt to complete the assessment. The teacher must also possess the interpretation skills necessary to analyze the assessment data collected from students. Inaccurate analysis of the data can lead to ineffective instruction, which results in missed learning opportunities for the student (Heritage et al., 2008).

Edvantia, a regional education laboratory founded in 1966 as a nonprofit education research and development corporation, partners with education practitioners as a service provider to assist in the advancement of student learning through professional development or consulting services. Edvantia recommends a clearly defined step-by-step process for the interpretation of assessment data. Due to the importance of properly interpreting assessment data to facilitate an

effective and efficient instructional process as well as feedback process, Edvantia defined an assessment interpretation process as part of their educational services professional development package. The assessment interpretation process lists these steps: (The Appalachia Educational Laboratory, 2005, p. 9):

1. Determine your purposes for analyzing the data.
2. Develop a question list.
3. Familiarize yourself with the data.
4. Organize the data.
5. Analyze the data.
6. Develop action steps.

Effective analysis of assessment data is an important step in the formative assessment process. It requires teachers to possess a separate set of sub skills to implement. Lewis (2006) also asserted the need for quality analysis and interpretation of assessment data and linked the analysis of data to content knowledge. In the presentation by Lewis (2006), she cited the co-director of CRESST at the 2005 CRESST conference "If you have lousy data, no matter how well massaged, it is not going to be useful" (p. 29). The data referred to in the interpretation and analysis process is directly related to the goals and objectives of the curriculum. Herman, Osmundson, Dai, Ringstaff, Timms (2011) asserted the importance of the teacher possessing a solid degree of content knowledge to carry out data analysis process:

> Because formative assessment is a dynamic process of evidence elicitation, analysis, and action, it clearly makes demands on teachers' content and pedagogical knowledge. Without such foundational knowledge, teachers'

> formative assessment may yield faulty decisions that could divert rather than promote student progress. (p. 2)

Providing Feedback to Students on Next Steps in Learning. Students have always been given feedback related to the school work they complete or assessments they take. Many individuals, including parents, students, and even professionals working in the field of education, agree that feedback to students comes in the form of grades (Black & Wiliam, 1998). Grades are one form of feedback provided to students and parents to communicate academic achievement. However, grades associated with communicating achievement levels on a completed unit of learning or an assessment is not the desired form of feedback cited by many education researchers or practitioners in the field to facilitate higher levels of student learning (Marzano, Pickering, & Pollock, 2001).

Other scholars have stated that grades are assigned by teachers to provide students with feedback on their level of proficiency (Black et al., 2004). Students most often utilize these assigned grades to rank themselves against other students rather than attempting to discern the degree to which they have truly mastered the desired goals or objectives. According to Black & Wiliam (1998), "research studies have shown if pupils are given only marks or grades, they do not benefit from the feedback on their work" (para. 39). Black and Wiliam (1998) viewed the utilization of grades as a form of feedback for students on their learning as serving only two functions. First, grades serve as a social function allowing a student to compare him or herself to other students. Second, grades serve as a managerial function, giving teachers a communication tool to inform students of their degree of mastery of the course goals and objectives. Assigning grades to students is such a complex (and sometimes controversial) issue that some educators have even proposed their abolition (Kohn, 1999; Marzano, 2000).

According to Marzano, Pickering & Pollock (2001), providing feedback to students related to specific levels of strengths and weaknesses concerning concepts or skills more effectively and efficiently improves student learning as compared with the practice of assigning a grade at the end of a unit. Feedback facilitated by the use of rubrics or set learning goals for specific skills or content knowledge if a powerful tool in the feedback process. Feedback needs to be specific in nature and easily understood (Marzano et al., 2001**)**

The feedback provided by the teacher is intended to facilitate a self-regulatory process impacting the student's cognition, motivation, and behavior (Nicol & Macfarlane-Dick, 2006). To activate the self-regulatory process in the student, the teacher begins by setting a task with clear goals or objectives. The teacher's actions are an external stimulus triggering an internal response by the student in the form of activating his or her prior knowledge related to the goal or task assigned. Student understanding of the learning goal or task aids in developing an understanding of the goal or task and sets in motion the tactics and strategies the student will utilize to accomplish the learning goal or task. A student's internal learning outcomes are then modified by the new understanding or skill sets created by applying tactics and strategies to accomplish the goals set by the teacher. The new internal learning outcomes of the student manifest in the form of an external product in the form of a completed project or paper, a completed test or quiz, a presentation, and so on. During this process, an internal feedback process is taking place in which the student is constantly monitoring his or her own progress against the stated goal. It is vital for the teacher to continue meeting with the student, provide feedback related to the student's progress toward the goal, and initiate further internal processing through additional questioning (Nicol & Macfarlane-Dick, 2006). If a teacher is to provide effective feedback to facilitate a student's self-regulatory process, the teacher must possess

content knowledge, the learning progressions in degrees of sophistication of said knowledge, and what that knowledge looks like in student performance (Heritage et al., 2008).

Building Teacher Competency through Effective Professional Development

Lewis (2006) contended that the response to high stakes state testing in the education community does not include enough professional development to increase teacher capacity and meet the growing demands mandated through NCLB. Lewis (2006) calls for professional learning aligned to the new learning standards. To create changes in teacher instructional practices, some educational researchers are calling for an increased focus on building the capacity of our country's teachers and tying it directly to the assessment reform agenda (Borko, 1997). Lewis (2006) reported the nation could continue to respond to the need for improved academic achievement in its schools by testing even more subjects in additional grades, but one of the changes needed was the real reform in focusing on professional development for all teachers across the United States.

In 2008, the Consortium for Policy Research in Education, in collaboration with five national research universities including Stanford University, University of Michigan, Harvard University, University of Pennsylvania, and the University of Wisconsin-Madison, founded a project called Strategic Management of Human Capital in Education (SMHC). The project's purpose was to review current research on effective practices to improve student achievement. The SMHC (2009) report asserted changes need to occur in teaching methodology or practice in order to improve student achievement. To introduce changes to teachers' methodology that have proven to be effective and lasting, practitioners need to adopt a model of professional development including an increased investment in time and money for teacher training similar to the heavy investment the business community puts into the capacity building of employees.

> The SMHC (2009) report also called for states to provide adequate funding for ongoing professional development, giving teachers time to collaborate on curriculum and instruction during the regular school day, additional days of training outside of the school day, and assistance and training from instructional coaches who are part of the school staff.

In the early 1990s, legislators in Missouri recognized the need to invest in the professional development of teachers and passed the Outstanding Schools Act SB287 (1993). This act set forth funding regulations mandating each school district in the state to spend 1% of the money received from the foundation formula to the professional development of teachers. Similar funding requirements are in place in most states across the country, calling for money to be set aside for the professional development of teachers. The training of new teachers offers sustained learning opportunities adjusted to the needs of each teacher as identified through their classroom experiences (Schleicher & Stewart, 2008).

According to a study conducted by Darling-Hammond and Cobb (1995), the commitment to the professional development of teachers in countries doing well on international student achievement tests is in stark contrast to the professional development provided to new teachers in the United States. The study looked at the professional development practices of twelve countries in the Asia-Pacific Economic Cooperation (APEC), and found that these twelve nations invested more time, support, and mentoring than teachers in the United States are provided.

High Quality Professional Development Defined

Quality professional development activities have been defined. If the professional development we provide our teachers is not effective in improving competency, the money and time spent would be a waste of valuable resources. High quality professional development (HQPD) and the activities associated with HQPD have been identified by the Missouri

Department of Elementary and Secondary Education (2006) and the National Association of Elementary Principals and National Staff Development Council (1995); HQPD is divided into three areas including process/context, activities, and topics.

All seven components of HQPD Part I are necessary to facilitate effective teacher learning in the utilization of formative assessment. Components 2, 3, and 4 of Part II are directly related to the facilitation of effective learning in the use of formative assessment, with all other components also relating significantly to the facilitation of teacher learning in the use of formative assessment. Components 1, 2, 4, 6, 9, and 10 in Part III are directly related in the facilitation of effective learning in the utilization of formative assessment, with all other components also relating significantly to the facilitation of teacher learning in the use of formative assessment**.** The Missouri Department of Elementary and Secondary Education (2006) criteria for HQPD closely aligned with that of the National Staff Development Council and the National Association of Elementary Principals (1995). These documents indicate that a definition of HQPD activities has been established along with the activities and topics supporting HQPD.

A Calling for Professional Development

Educational researchers Black and Wiliam (1998) reviewed and synthesized over 250 articles on the topic of formative assessment and the impact of assessment on student learning. Their report indicated formative assessment improved student achievement for all students and proved to be significantly more effective in improving achievement of low performing students. Out of the 250 articles reviewed, Black and Wiliam (1998) considered several dozen of the articles to be of sufficient academic rigor with adequate experimental controls allowing for reasonable conclusions to be made based upon the positive effects of one half to one full standard deviation. Stiggins (2002) stated:

> Hypothetically, if assessment for learning, as described above, became standard practice only in classrooms of low-achieving, low-socioeconomic-status students, the achievement gaps that trouble us so deeply today would be erased. I know of no other school improvement innovation that can claim effects of this nature or size. This result has direct implications for districts seeking to reduce achievement gaps between minorities and other students. (para. 43)

Bloom (1975) provided additional support for the use of classroom assessment as an instructional tool to inform and drive instruction in a study in which he accounted for improved student achievement in an experimental group attributable to the use of formative assessment. Further evidence exists in the work of Wylie et al. (2009) indicating that professional development for educators needs to address formative assessment as a way to improve both content and process.

Based upon the research in related literature, it appears reasonable to call for the use of professional development to increase teacher capacity in the skills necessary to effectively and efficiently implement formative assessment and the associated instructional strategies in the classroom.

Classroom Instruction Strategies Linked to Assessment

Formative assessment Strategies to improve student achievement need to be clearly defined. A report outlining six principles to facilitate improvement in student achievement was published (SMHC, 2009) stating using formative assessment was one of the six key principles for ensuring teachers are effective. Teachers must work in collaborative teams to analyze student data, tailor instruction to diverse student needs, deliver content in multiple ways, and quickly measure student progress in addition to their own progress (SMHC, 2009). Each of these strategies is

reliant upon the teacher's content knowledge and the learning goals established. According to Bell & Cowie (2001), formative assessment is the process by which assessment is utilized to provide feedback to the teacher on the effectiveness of lessons taught and provide feedback to the students to further their understanding of concepts and improve their skills. The teacher's degree of success in the implementation of each one of these steps is reliant upon the teacher's degree of content knowledge.

Dunn and Mulvenon asserted (2009) the intent of NCLB was to move teachers toward the utilization of data to inform their instruction and in turn begin to make systemic improvements to instruction. The goal is to help teachers better utilize formative assessment. However, this tool to improve student learning is not utilized effectively in classrooms due to the fear surrounding the high stakes testing of students. Dunn and Mulvenon (2009) contend that teachers' fear of assessment is rooted in the utilization of high stakes testing for NCLB and the lack of knowledge on how to utilize these high stakes assessment to aid in improving instruction:

> Many teachers do not feel empowered when dealing with assessment issues as there is a glaring absence of understanding in both the classroom and the literature with regard to how to fully use the power of both summative and formative assessments in education...It is important for teachers, administrators, researchers, and policy makers to share a common language related to assessment so as to unlock the power of assessment and create positive changes in student achievement. (p. 3)

In addition to creating a common language to implement formative assessment strategies to drive instruction, Cobb, McClain, Lamberg, and Dean (2003) called for the professional development process to be administered locally, sustained over time, and involve collective

participation by all related team/grade/school members to allow for sensitivity to local needs. Cobb et al. (2003) also referenced researchers (e.g., Cohen & Hill, 2000; Garet, Birman, Porter, Desimone, & Herman, 1999) calling for professional development to be ongoing and local to support teachers' instruction. Wylie, Goe, & Educational Testing (2009) argued the current Professional Learning Community movement attends to the importance of time, a local school-based approach, and a collective action taken by all stakeholders. Further, Thompson et al. (2009) reviewed the work of several researchers citing very similar conclusions to the information presented in the SMHC report (2009) on the use of effective professional development to make sustainable changes to teaching methodology and practice.

These theories on effective professional development are not new and apply to learning among adults in and out of the education field. In a study on how Japanese companies encourage innovation, Nonaka and Takeuchi (1995) cited some of the very same components in the creation of knowledge. They assert that, for knowledge to become "solid," it must be acquired through doing. In order for new knowledge to become operational, it must be linked to existing knowledge and then internalized to create a new structure whereby the individual can readily access the new tacit knowledge in a manner that is meaningful and useful. The assertion is that one must connect the new knowledge with prior knowledge, be given the opportunity to practice the new knowledge and then reflect upon the ramifications of attaining the new knowledge, and the knowledge must be made accessible to accomplish a given task (Nonaka and Takeuchi 1995).

High Quality Professional Development

According to Richards (2009) and Thompson et al. (2009), practitioners in the field of education have failed to recognize that most of the professional development taking place does not include the necessary components required in learning new theory or ideas, time to process,

practice, reflect, and make adjustments. Research indicates that a good understanding exist of methods for supporting teachers through effective and efficient professional development to change pedagogical methods improving student achievement. The key is to utilize an - understanding of effective professional development to affect reform in instructional practices in the classroom.

Educators should begin to focus on developing the individual competencies necessary to implement formative assessment instructional strategies.

A group of researchers and practitioners (e.g. DuFour, DuFour, Eaker, & Many, 2006; Marzano, 2009; Wiggins & McTighe, 2011) asserted a necessary first step for the effective implementation of formative assessment strategies. Teachers must possess the ability determine the learning goals and objectives of the curricular content they are responsible to teach. To facilitate the development of content learning goals and objectives teachers must have a solid base of content knowledge. It is difficult at best to determine learning goals if knowledge of the content does not exist. DuFour, DuFour, Eaker, and Many (2006) referred to learning goals and objectives as information teachers need to determine related to their first of three key questions what is it that we want our students to learn. This is the first step in formative assessment process outlined in his Professional Learning Communities instructional process. Marzano (2009) defined goals and objectives as something we want a student to be able to know or do. To be a goal or objective it must meet one of three separate conditions including performance, conditions and criteria. Marzano (2009) also asserted the necessity to determine goals and objectives to facilitate development and delivery of instructions including lesson design, assessment, determining next steps and feedback. Wiggins and McTighe (2011) framed learning goals and objectives through their Understanding by Design approach. In this approach teachers need to

develop enduring understandings, essential questions and the goals and objectives aligned to enduring understandings and essential goals.

Summary

The importance of formative assessment instructional strategies has been addressed in literature by researchers and education professionals. Assessing students requires a determination of what we want students to learn (DuFour & Eaker, 1998). The research indicated determining what we want students to learn had been addressed as early as the forties by Tyler (Guskey, 1994). Issues have been identified making the development of learning goals and objectives difficult. Black and Wiliam (2010) argued, "Teachers will not take up ideas that sound attractive, no matter how extensive the research base, if the ideas are presented as general principles that leave the task of translating them into everyday practice entirely up to the teachers" (p.87). It is important to support our teachers in learning each skill necessary to be effective in their instruction.

Chapter III

Methodology

Educational research studies indicate that strategies for determining learning goals vary greatly across the country. As well, research supports the position failing to provide our teachers time and support necessary to develop essential learning goals presents a critical problem in classroom practice (DuFour, DuFour, Eaker, & Many, 2006). Developing essential learning goals is not an easy task, and teachers report they are uncomfortable developing quality essential learning goals on their own. They are aware of the benefits of taking the time to determine exactly what they want their student to learn, and many report having attended some form of professional development addressing this issue**.** (DuFour, 2011)

Research Question

Research has established the importance of teachers' ability to determine the essential learning goals of the content they are responsible to teach and the issues around lack of support for teachers' planning and professional development time. Thus, this study is designed to answer the following research question: Will the development of essential learning goals result in improved student achievement?

Alternative Hypothesis

In order to obtain data related to the stated problem and research question, the following alternative hypothesis was developed:

H_1: The development of essential learning goals in grades two through five in the area of communication arts will impact student achievement.

Null Hypothesis

In order to obtain data related to the stated problem and research question, the following null hypothesis was developed:

H_o: The development of essential learning goals in grades two through five in the area of communication arts has no impact on student achievement.

The alternative hypothesis addresses the idea that a relationship exists between the development of essential learning goals and student achievement. According to the null hypothesis a relationship may not exist between the development of essential learning goals and student achievement, but an inability exists to rule out other factors which may affect the outcomes of the study.

Population of the Study

The sample population of 19 teacher participants in this study works at two middle schools in a middle class suburban school district near a major metropolitan area. Nine teacher participants compose the sample from School (A_t)—the school with teachers receiving the treatment--ten teacher participants compose the sample from School (B_{nt})—the school with teachers not receiving the treatment. In the group receiving the treatment, population was based on proximity to the researcher. Teachers in the control group were selected based on four factors. First, the demographic characteristics of that sample are similar to the treatment group sample. Second, both the treatment group and control group receive a similar amount of professional development time. Third, the population in the control group had not participated in professional development activities related to creating essential learning goals. Fourth, student archival data of the school district for both groups is available to the researcher.

All teachers were employed in a fully accredited, suburban school district with nearly

12,000 students (DESE, 2011). Additional members of the population's demographic information includes average years of teaching experience, percentage of teachers who earned a master's in education degree, and the number of male and female participants (see Table 1 on page 121; DESE, 2011).

The study reviews student achievement scores in the area of communication arts for grades two through five with a total population of 808 students from the two schools during the 2011-2012 school year. School A_t has population of 318 students enrolled, and School B_{nt} has a population 490 students enrolled. Students were selected if they had archival data available from the school district and attended in grades two through five during the 2010-2011 and 2011-2012 school years. Student demographic information also includes the following: the number of students attending each school; the number of males and females attending each school; the ethnic origin of students as reported in percentages in Annual Yearly Progress (AYP) data; the percentage of students qualifying for free or reduced lunch; the student attendance rate for each school; and the number of disciplinary actions resulting in suspension or expulsion (See Table 2 on page 122; DESE, 2011).

The student population at both School (A_t) and School (B_{nt}) was assigned to classes with the following parameters: approximately the same number of male and female students; the special education caseload presents roughly the same degree of difficulty to the classroom teacher; a relatively equal number of non-special education students with low, middle, and high academic skills; and a relatively equal number of non-special education students with low, moderate, or severe behavioral/emotional concerns.

Instruments

Student measure instrument. The Edison Schools E-valuate Benchmark Assessment

Tool, designed to measure students' academic progress in the areas of Mathematics and Communications Arts, served as the instrument to obtain student academic achievement scores. The tool is available for use by schools to measure student performance in Grades two through eight. For the purpose of this study, the Communication Arts assessment for Grades two to five will be utilized. This tool was chosen due to its archival nature, with data available for both School A_t and School B_{nt}. The archival data from this tool exist only for grades two through five.

The E-valuate tool was administered monthly using the same procedures at both schools. The study utilized archival data from the 2010 -201 and 2011 – 2012 school years. The E-valuate assessment was used as a formative assessment to inform instruction, develop remediation groups, and predict success on the state-level end-of-the-year exams. The questions are designed to assess grade level expectations in end-of-the-year state tests aligned to cover the content and process goals of the state in which the school exists. Each assessment is made up of 20 questions. Each of the twenty questions is worth 5 points, for a total of 100 points possible. Students are not provided any feedback as to their performance on the test while it is being is taken. Students' results are reported in the following ways: total percentage of points earned; item analysis of each question based upon correct or incorrect; and number of items correct or incorrect based upon curriculum content benchmarks.

The usual test taking procedure is that students take the Communication Arts and Mathematics assessments each month. Students take a paper version in their classroom following the test-taking procedures established for all classroom assessments. Students are spaced apart to keep students from observing other students' answers, are to complete all questions on the assessment, and are required to remain quiet while the testing is taking place. Upon completion of the test, students are to review their answers for accuracy and completeness.

After the students complete their E-valuate assessment on paper, the students move to the computer lab to enter their responses using the web-based instrument. All test-taking procedures are followed in the lab while students are entering their responses. The validity, reliability, and item difficulty information is provided by the assessment development company Edison Learning (see Appendix D on page 96 for complete validity and reliability information).

Teacher Measure Instrument. An observation protocol was developed by the researcher to determine if teachers were using essential learning goals in the classroom. A teacher was given a "0" when the teacher was not observed addressing essential learning goals in the classroom or a value of "1" if observed using essential learning goals in the classroom. The assignment of a value of one or zero is necessary for calculations in the repeated measures ANOVA. The researcher made a total of five nonscheduled observations to look for evidence (indicators) of the use of essential learning goals in the classroom. Indicators consisted of: appropriate essential learning goals posted in the classroom, essential learning goals addressed in the teacher plan book, or essential goals addressed in the lesson. In order for an observation to be assigned a value of "1", two of the three indicators needed to be present in the observation **(**see appendix E on page 101 for observation form). A teacher was assigned a "1" if observed using essential learning goals in three of the five observation trials (see Appendix F on page 102 for form to assign final value).

Statistical Analysis

Analysis of variance between control and experiment groups. The sample in this study, both students and teachers, were subsets of their true populations. Inferential statistical methods were employed due to the desire to make claims concerning the population (DeCaro, 2003).

Exploratory data analyses were performed to assure the variables met the assumptions underlying the ANOVA and regression tools. Measures of central tendency and dispersion were reported as appropriate for the type of variable and, combined with graphic distribution analysis; confirmed normal distribution of all variables. If any skewness or kurtosis issues arose, transformation was applied prior to inclusion of the variable in the analysis. As regression analysis was employed, attentions to any outliers in variables were paid on both the bi- and multi-variate levels. Equality of variance amongst the variables was tested with Box's M test as appropriate. Collinearity was assessed using both graphic (scatter plot) and statistical (correlation matrix) analysis. Finally, the assumptions of normal distribution of the residuals in any regression analysis were confirmed in the process of analyzing the model.

In order to answer the research question, data analysis focused on examining the variance between the control and experimental groups along with the variance created by differences existing among the participants of the experimental group using the ANOVA family of tools. A series of 14 data points were utilized to collect student achievement information for analysis in the presence of teacher and group (= school) assignment.

The investigation of a possible treatment effect resulted in a between group design (Furlong, Lovelace, & Lovelace, 2000). The use of this design helped to determine if the treatment created a difference**,** as compared with the control group, leading to between group variability. The ANOVA calculations were used in order to identify variations between the control and experimental groups as response to the treatment or chance.

Addressing variability among teacher population. To address the variability existing among the teacher participants within the experimental group, a multivariate linear model was utilized. This analysis addressed the differences existing among the teacher participants of the

experimental group and the multiple relationships existing between the variances of the experiment participants and achievement. The three variability factors considered for the participants in the experimental group were years of teaching experience, years of education beyond a bachelors, and observations of the use of essential learning goals in the classroom. The variables experience and years of education were continuous variables and the observed use of essential learning goals in classroom instruction was dichotomous.

An analysis of the variability among the experimental group participants allowed for overall variability among the participants to be identified as main effects of the participants and the interactions of the participants' differences (Furlong et al., 2000). This analysis was used to systematically determine the effect, if any, on achievement related to each of the three independent variables. It also allowed for analyzing the effects, if any, on the interactions of the first, second, and third, independent variables on achievement (Furlong et al., 2000). A repeated measure ANOVA was utilized to investigate variance within the experimental group subjects.

Analysis software package. For the purpose of setting up a spreadsheet and making the calculations associated with the statistical methods utilized in this study, the SPSS (© IBM) software program was utilized. This software package performs general linear and mixed model procedures that are necessary for the statistical methods chosen to utilize in the analysis of data. This SPSS package also has the capabilities to make calculations using binary code, which was utilized to address the within subject variability of the participants in the experimental group. The SPSS (© IBM) package also deals with multivariate data embedded in within subject variances existing among the experimental group participants

Professional Development Treatment

Literacy coach description. A literacy coach facilitated the professional development

for the treatment group in the study and was assigned to the building by the school district central office. The literacy coach was placed to meet the call for job embedded approaches as suggested by the National Staff Development Council (NSDC, 2001). The literacy coach assigned to the school study site meets many of the desired qualifications established by the NSDC (2001). The literacy coach earned a Master's Degree in Literacy, has had successful teaching experience at the elementary grade level, and has experience working with teachers in a collaborative environment to improve student achievement. The literacy coach served as a classroom teacher for fifteen years, has excellent presenter skills, experience modeling lessons, and classroom observations skills. She has worked at the study site for three years supporting teachers in the area of reading instruction.

Treatment description. The professional development treatment was designed to meet high quality professional development process, content, and activity standards established by the Missouri Department of Elementary and Secondary Education (2006), the National Staff Development Council, and the National Association of Elementary Principals (1995).

Professional development activities, led by the literacy coach, were provided to the experimental group from August 2011 to April 2012. Participating teachers were guided through the process of developing essential learning goals. Experimental group participants engaged in a book study group reading *The Understanding by Design Guide to Creating High Quality Units* by Wiggins and McTighe (2011). Participants discussed enduring understandings, essential questions, learning goals and objectives, and learning progressions. When the participants reported an understanding of each of the components, they initiated the work of developing each component of the essential learning goals of the curriculum they are responsible to teach.

Participants developed an agreed upon set of curriculum enduring understandings by

grade level teams in breakout sessions. Each grade level team presented their developed enduring understandings, identified common understandings, identified differing understandings, and reached a consensus on the enduring understandings for the entire group of participants. The process for the development of essential questions followed the same procedures as those utilized to develop enduring understandings and used the *Understanding by Design Guide to Creating High-Quality Units* book. In addition, essential questions were collected from local school districts by the literacy coach in the development of essential questions.

Upon completion of the enduring understandings and essential questions, study participants aligned their enduring understandings and essential questions with the topics/skills of their respective Communication Arts grade level curriculum. Upon completion of this activity, all participants met to report their work, take suggestions, and refine their enduring understandings. Participants also worked to assure vertical alignment existed for all topics and skills in Grades two through five. Each grade level team developed their alignment of enduring understanding and essential questions to the topics/skills of their curriculum using a template design created by the literacy coach. The following is an example.

Enduring Understanding 1	**Enduring Understanding 2**	**Enduring Understanding 3**
Essential Question 1	**Essential Question 1**	**Essential Question 2**
Topic/Skills1	**Topic/Skills 1**	**Topic/Skills 3**
Topic/Skills 2	**Topic/Skills 3**	**Topic/Skills 4**
Essential Question 2	**Essential Question 2**	**Essential Question 3**
Topic/Skills1	**Topic/Skills 4**	**Topic/Skills 5**
Topic/Skills 5	**Topic/Skills 7**	**Topic/Skills 9**
Essential Question 4	**Essential Question 4**	**Essential Question 4**

Figure 1. Method for organizing enduring understanding and essential questions to topic/skills.

After completion of the alignment template participants determined the curricular goals and objectives of the participants' curriculum related to the identified topics and skills. The study

participants worked through this task supported by the literacy coach and the text *Designing and Teaching Learning Goals and Objectives: Classroom Strategies that Work* (Marzano, 2009). Teachers focused on four elements of theory presented in Marzano's text including: the two dimensions, different types and communication of learning goals; along with the importance of feedback in the instruction process. Study participants also determined the performance, conditions, and criteria of their goals and objectives as well as the specificity and difficulty of their learning goals. Participants also identified whether a goal was related to mastery or performance. This work was important because it directly impacted goal development, communication, and the feedback provided to students (Marzano, 2009). Wiggins and McTighe (2011) asserted the importance of teachers making a deliberate effort to assure that their learning goals and objectives are directly aligned with topics and skills flowing back up to the essential questions and finally back to the enduring understandings.

Participants studied the theories of Hess (2007) on identifying learning progressions for specific goals or skills. Participants discussed the four interrelated guiding principles of learning progressions, questions for the development, refinement, and validation of learning progressions, and strategies for beginning the development of learning progressions. Utilizing these concepts enabled the participants to focus on the "smallest grain sizes" (Hess, 2007, p. 13) of learning concepts necessary to inform/plan instruction, identify breakdowns in student understanding, and provide feedback to students. Each grade level team developed a learning progression for each topic or skill identified (see Appendix G on page 103 for an example of the learning progression template).

Upon completion of the development of the learning progression, each grade level team presented their completed templates, identified common components, identified differing

components, addressed vertical alignment, and determined the next steps to further incorporate the use of essential learning goals into daily instruction.

To support the implementation of the newly developed essential learning goals in daily instruction, the literacy coach utilized a learning lab approach (Sweeney, 2007), allowing the teachers to work in a supportive, sustained, and collaborative environment related directly to their daily instruction. All of these components of the learning lab approach are considered to be essential for high quality professional development (Missouri Department of Elementary and Secondary Education, 2006; The National Staff Development Council and the National Association of Elementary Principals, 1995).

Issues addressed on the learning lab observation sheet address questions such as: On which essential learning goals in the instruction component are the observers to direct their focus? What is the teacher's responsibility in the lab? What are the responsibilities of the observers in the lab? After the observation, the participants of the lab will meet and review their observations led by a set of prompts such as: "What went well…? I have these questions…; and Have you thought of…? After these conversations, the participants determine the next steps for the participant observed and the participants observing (see Appendix H on page 104for an example of learning lab observation sheet).

Summary

This study addressed the effect of professional development in establishing essential learning goals on student achievement. The research question was: If teachers were supported with time and professional development to improve their competency in the development of essential learning goals in grades two through five in the area of communication arts, would the development result in improved student achievement? The alternative hypothesis was that the

development of essential learning goals in grades two through five in the area of communication arts would impact student achievement. The null hypothesis was that the development of essential learning goals in grades two through five in the area of communication arts has no impact on student achievement. Nineteen teachers and 597 students participated in the study. Data on the teacher participants was collected using two observation tools and were analyzed using a multivariate linear regression model to determine variance within three independent variables related to the teachers in the experiment group. Archival data of the students was utilized in a general linear regression model to determine variance between groups' student achievement.

Chapter IV

Results and Analysis

This chapter presents the data collected to address the stated research questions and hypotheses discussed in chapters 1 and 3. This chapter is organized as follows: Introduction, descriptive statistics, presentation of data and analysis, and a summary.

Introduction

The purpose of this study was to determine weather or not the development of essential learning goals in grade two through five would result in improved student achievement in communication arts reading scores. Results indicated an effect occurred in the student reading scores of the experiment group in the post year of the study. The hypotheses related to the study question were as follows:

Alternative hypothesis. H_1: The development of essential learning goals in grades two through five in the area of communication arts will impact student achievement.

Null hypothesis. H_o: The development of essential learning goals in grades two through five in the area of communication arts has no impact on student achievement.

The study addressed the application of a year-long professional development intervention and impact on a group of nine teachers as well as student achievement scores in communication arts. The achievement data of the students in the classrooms of the teachers engaged in the professional development treatment were compared to the achievement data of students in classrooms whose teachers were not engaged in the professional development treatment. There were ten teachers in the comparison group not engaged in the professional development treatment. The year-long professional development treatment was designed to facilitate the acquisition of a lesson planning strategy utilizing the development of essential learning goals in

the area of reading instruction.

Data for the baseline year were established utilizing communication arts student achievement scores from the 2010-2011 school year. Data for the post analysis was established utilizing communication arts scores from the 2011-2012 school year. Student achievement data were generated by administering an end of year formative assessment tool developed by the Edison Learning company. The formative assessment tool is called the E-valuate Assessment of Communication Arts. The E-valuate formative assessment tool can be administered to students in grades two through eight. The assessment was administered to the students monthly from September through March.

In addition to addressing the possible effect of the professional development treatment, the study addressed additional covariant factors including years of experience teaching, years of education beyond a bachelor's degree, and the observation of teachers in the treatment group using essential learning goals within their instructional methods. Factors involved in the observation of the use of essential learning goals included goals stated in lesson plans, goals posted in the classroom, and the goals stated during the lesson.

Within the presentation of the data and analysis section, analyses were conducted in two different formats. The first analysis looked at two groups across fourteen different data points (7 baseline, 7 post) over a two year period of time (year 1 = baseline; year 2 = post). A Follow-up analysis looked at two groups across seven data points over a year-long baseline period of time. Additionally, the follow-up analysis looked at two groups across seven data points over a year-long post period. The follow-up analysis was conducted using a larger sample size of cases to confirm the results of the first analysis. The data and analysis section will be presented in the following sequence: ANOVA (2 x 14, N = 593, n= 225), descriptive statistics, multivariate test

(Wilks' Lambda), Mauchlys Test of Sphericity, corrected test of within-subjects effects (Greenhouse-Geisser) test of within-subjects contrasts, and a test of between subjects effects. An ANOVA (2 x 7, N = 318, n = 85) for the post year, goal use analysis in A_t only. An ANCOVA for (2 x 14, N = 593, n= 225) for both teacher experience and education beyond a bachelor's degree, and an ANOVA, (2 x 7, N=593, n = 407) post year only, using a larger number of cases to confirm results of effect in post year of first ANOVA analysis.

Descriptive Statistics 2 x 14 Analysis

The main analysis was conducted looking at two groups of students over a two-year period of time. Over the two-year period of time, there was a total of fourteen data points representing fourteen different assessments administered to students in grades 2 through 5 at both schools (September through March, baseline; September through March, post).

There were a total of 593 (*N*=593) total cases possible for the study when all students were included in the attendance of the 19 teachers instructing students in grades 2 through 5 in both the comparison and treatment groups. Four cases (.67% of total cases) were removed from the initial dataset due to the students moving between schools causing their data to exist in both the comparison and treatment group's data. In addition, two cases (.34% of total cases) were removed due to the cases missing two or more scores out of the seven possible data points in a given year.

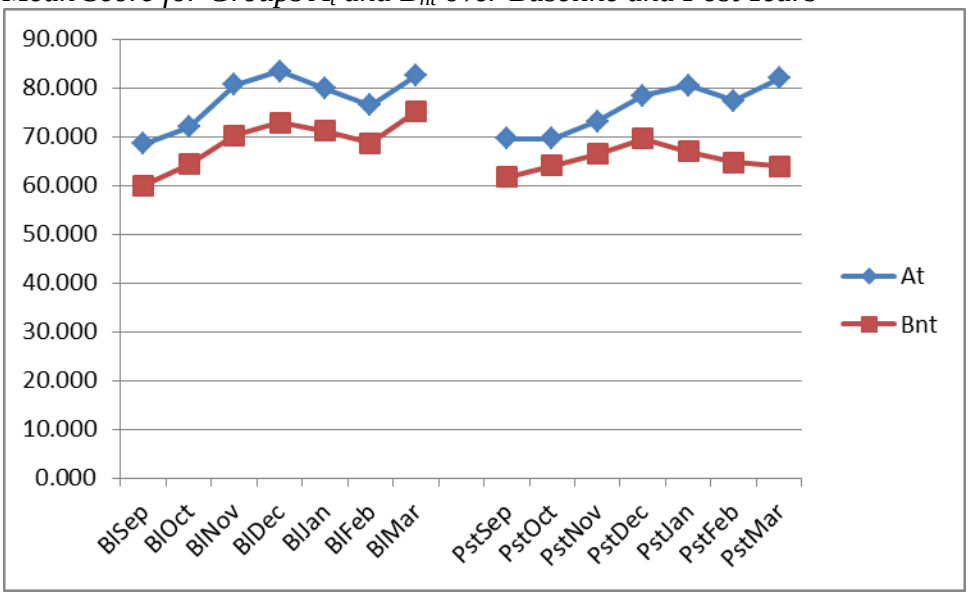
Three hundred sixty-eight cases (62.05%) were excluded from the first ANOVA (2 x 14) because the student was not represented in both the baseline and post periods. Specifically, of the 411 students with complete baseline data, 182 did not have post period data (e.g., because they were in sixth grade during the post period, or transferred to another school, etc.); of the 407 students with post period data, 186 did not have baseline data (e.g., because they were in first

grade during the baseline period, etc.). This resulted in a total of 225 (n = 225) valid cases containing both a baseline and posttest set of test scores. The cases were categorized into n = 85 for group A_t and n = 140 for group B_{nt}. The remaining total of 225 cases resulted in a total of 3150 data points for analysis of the baseline and post years.

In the baseline year, the mean score for group A_t increased from 68.80 to 82.62. The mean score for group B_{nt} increased from 59.97 to 75.06. During the post year, the mean score for group A_t grew from 69.62 to 82.19. The mean score for group B_{nt} increased from 61.72 to 63.99 (see table 3 on page 123 for complete information on mean scores). The mean scores for both groups behaved as expected. Both groups' mean scores increased over time from September through March indicating that learning of end of year expectations had taken place. The increases of mean scores occurred for both the baseline and post years. However, both groups' mean scores did not increase by the same degree during the post year. Graph 1 displays the relationship of the mean scores for grades two, three, four, and five over the two year period of time.

Graph 1

Mean Score for Groups A_t and B_{nt} over Baseline and Post Years



The relationship between the two groups' mean scores is parallel until December of the post year. At the December data point the two mean scores begin to separate with the distance of separation increasing through the March data point. There are a variety of possible causes for the change in the relationship of the two groups mean scores. No conclusions can be drawn from the descriptive statistics.

Analysis ANOVA 2 x 14 Baseline and Post

A group of multivariate tests was run to analyze the data from both the baseline and post years. The Wilks' Lambda test was utilized as it is the test most often used to identify multivariate differences in mean scores between and within groups (Everitt and Dunn, 1991). The effect of learning over time (i.e., main effect of time) was supported in this test. The Wilks' Lambda test resulted in $\lambda = .383$, $F(13, 211) = 26.101$, $p = .000$, $\eta^2 = .617$. The observed power of the test was 1.000. Again, learning was expected to take place over time. The assessment tool measures learning over time of the grade level expectations. The Partial Eta Squared value indicates that 61.7 percent of the change in scores could be attributed to learning over time. A Wilks' Lambda test was also evaluated regarding the Time x Group interaction effect. The Wilks' Lambda test for Time x Group interaction resulted in $\lambda = .850$, $F(13, 211) = 2.684$, $p = .001$, $\eta^2 = .150$. The observed power of this test was .991. With the F_{crit} value being greater than one and a significance value of .001 there is a significant effect associated with differential change in the study groups over time (Furlong, Lovelace, & Lovelace, 2000). The Partial Eta Lambda also indicates a significant effect with $\eta^2 = .150$ (Levine and Hullett, 2002).

A Mauchly's Test of Sphericity was run to test the repeated measures analysis of variance assumption of sphericity. Basically, sphericity is satisfied when the variances of the differences between all possible pairs of the levels of a repeated measures variable are equal. More

technically, this test was conducted to test for inflated p-values for the F statistics due in part to the error covariance matrix of the orthonomalized transformed dependent variables being disproportional relative to an identity matrix (1s on the diagonal, 0s elsewhere). The Mauchly's Test of Sphericity resulted in $X^2(90) = 1354.808$, $p = .000$, $\dot{\varepsilon} = .338$. Since Mauchly's Test is highly sensitive to even minor deviations from sphericity, p-values of less than .001 (as indicated here) are generally interpreted as significant violations of the assumption. Thus, the test indicated that the assumption of sphericity was violated. In these situations, adjustments to the degrees of freedom for the multivariate tests are made, and the Greenhouse-Geisser adjustment was made here to the Time x Group interaction effect. This adjustment produces robust F and p-values despite the sphericity violation. The tests resulted in $SS_x = 7921.519$, $F (1802.330, 4.395) = 2.285$, $p= .052$, $\eta^2 = .10$ and the observed power was .700. Thus, despite the violation of sphericity, the Time x Group interaction was retained as significant.

A test of within-subjects contrasts (13 single degree-of-freedom polynomial contrasts represent the Time x Group interaction effect) was also completed. The Time x Group interaction was significant for both quadratic trend (ESS =3333.765, $F (1, 223) = 13.899$, $p = .000$, $\eta^2 = .059$, observed power = .960), as evidenced in the B_{nt} trend in Graph 1, and cubic trend (ESS = 1464.035, $F (1, 223) = 4.391$, $p = .037$, $\eta^2 = .019$, observed power = .550), as evidenced in the A_t trend in Graph 1. The mean values increased during the baseline year, decreased after the summer break, and increased again during the post year in the A_t group. In the B_{nt} group, however, mean scores did not increase over time during the post period, such that the baseline differences between the A_t group and the B_{nt} group increased significantly during the January-March phase of the post period. The within-subjects contrasts tests support the findings utilizing the multivariate Wilks' Lambda.

Descriptive Statistics and ANOVA for Goal Use in the A_t (Intervention) Group.

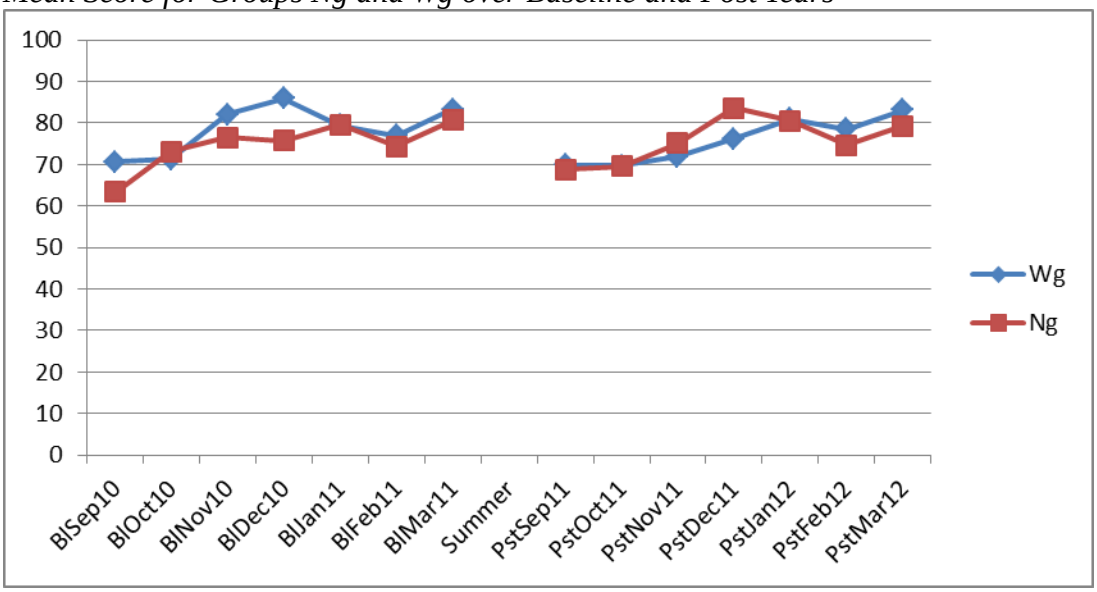
In the A_t group, over the two year period including the baseline and post years the number of cases out of the total cases of 593 that included both a pre and post score was 85, $n = 85$. The number of cases engaged with teachers using goals in the post year was 63. The number of cases engaged with teachers not using goals in the post year was 22. A teacher was assigned a value of 1 if observed utilizing essential learning goals in their lesson planning and instruction. A teacher not observed utilizing essential learning goals in their lesson planning and instruction was assigned a value of 0. To be assigned a value of 1 for the post year, a teacher was observed at 5 different points. For each observation point a teacher needed to be observed addressing essential goals in their planning, goals needed to be posted in the classroom and the teacher needed to be observed referring to the goals during instruction. Of the 5 observation points over the post year a teacher needed to be assigned a value of 1 in 3 of the 5 observation points (see Appendix J on p. 106 for observation scores).

In the baseline year, the mean score for students of teachers with goal use (WG) in the post year increased from 70.68 to 83.30. The mean score for students of teachers without goal use (NG) in the post year increased from 63.41 to 80.68. During the post year, the mean score for group WG grew from 69.92 to 83.22. The mean score for group NG increased from 68.77 to 79.23 (see Table 3 on page 123 for complete information on mean scores). Thus, the Time x Post Year Goal Use interaction effect was not significant, $\lambda = .752$, $F(13, 71) = 1.798$, $p = .06$. With correction for violation of the sphericity assumption, the p-value for this interaction was increased to well beyond established levels of significance: $p = .157$. Graph 2 clearly demonstrates the lack of an interaction effect attributable to post year goal use in the A_t group teachers. Thus, the explicit use of goals (e.g., posting them in the classroom) did not have an

effect on student reading scores beyond the effect of the professional development intervention itself in the A_t group.

Graph 2

Mean Score for Groups Ng and Wg over Baseline and Post Years



Analysis ANCOVA 2x14 with Teacher Experience as Covariate

Teacher experience was significantly higher in the At group (M = 12.2, SD = 9.2) relative to the B_{nt} group (M = 10.1, SD = 4.6), t (413) = 3.1, p = .002 (this comparison remained significant even after adjusting for unequal variances between groups, t (251.9) = 2.9, p = .004). As a result, a 2 x 14 ANCOVA was calculated with teacher experience as a covariate in the model. The primary effect, Time x Group interaction, remained significant with experience included in the model, $\lambda = .875$, F (13, 210) = 2.297, p = .007, although the effect size was smaller, $\eta^2 = .125$, relative to the unadjusted model presented above. Including experience in the model was analogous to comparing changes in reading scores over time if the two groups, A_t and B_{nt}, were equivalent on experience. Predictably, although the Time x Group interaction remained significant, the mean scores in the At group were adjusted downward, while the scores in the B_{nt}

group were adjusted upward. This effect of the covariate served to diminish, but not eliminate, the January-March differences between the two groups in the post year. The means and standard errors, adjusted for teacher experience, appear in Table 4 below.

Table 4

Adjusted Means and Standard Error for Teacher Experience

Group	Time	Covariate Adjusted Mean	Covariate Adjusted Standard Error
A_t	BL-Sep	67.566	2.068
	BL-Oct	71.833	2.052
	BL-Nov	80.759	2.249
	BL-Dec	83.16	2.061
	BL-Jan	80.031	2.023
	BL-Feb	75.36	1.928
	BL-Mar	81.608	1.898
	PST-Sep	69.007	2.204
	PST-Oct	68.932	2.169
	PST-Nov	71.921	2.220
	PST-Dec	78.525	2.447
	PST-Jan	79.84	2.279
	PST-Feb	76.926	2.258
	PST-Mar	80.546	2.273
B_{nt}	BL-Sep	60.721	1.587
	BL-Oct	64.366	1.574
	BL-Nov	70.154	1.726
	BL-Dec	72.967	1.581
	BL-Jan	70.846	1.552
	BL-Feb	69.16	1.479
	BL-Mar	75.674	1.456
	PST-Sep	62.089	1.691
	PST-Oct	64.499	1.664
	PST-Nov	66.955	1.703
	PST-Dec	69.36	1.877
	PST-Jan	67.511	1.749
	PST-Feb	65.03	1.732
	PST-Mar	64.99	1.744

Analysis ANCOVA 2x14 with Years of Education beyond a Bachelor's Degree as Covariate

Conversely, teacher years of education beyond a Bachelor's degree experience was significantly lower in the A_t group (M = 2.6, SD = 1.5) relative to the B_{nt} group (M = 3.7, SD = 1.9), t (413) = -6.1, p = .000 (no adjustment for variance inequality was needed for this comparison). As a result, a 2 x 14 ANCOVA was calculated with years of education as a covariate in the model. As in the previous ANCOVA, the primary effect, Time x Group interaction, remained significant with education years included in the model, $\lambda = .845$, F (13, 210) = 2.964, p = .001, and the effect size was not fundamentally changed, $\eta^2 = .155$, relative to the unadjusted model presented above. Again, including education years in the model was analogous to comparing changes in reading scores over time if the two groups, A_t and B_{nt}, were equivalent on education years. The slight effect of the covariate served to adjust upward the mean scores in the At group and adjust downward the mean scores in the Bnt group, but without appreciably changing the Time x Group interaction. The means and standard errors, adjusted for education years, appear in Table 5 below.

Table 5

Adjusted Means and Standard Error for Education in Years

Group	Time	Covariate Adjusted Mean	Covariate Adjusted Standard Error
A_t	BL-Sep	69.561	2.136
	BL-Oct	72.313	2.100
	BL-Nov	80.201	2.304
	BL-Dec	81.826	2.095
	BL-Jan	78.092	2.062
	BL-Feb	76.701	1.987
	BL-Mar	83.427	1.954
	PST-Sep	69.783	2.264

	PST-Oct	69.795	2.231
	PST-Nov	73.566	2.278
	PST-Dec	78.973	2.504
	PST-Jan	82.232	2.334
	PST-Feb	76.967	2.315
	PST-Mar	82.865	2.360
B_{nt}	BL-Sep	59.509	1.624
	BL-Oct	64.074	1.597
	BL-Nov	70.492	1.752
	BL-Dec	73.777	1.593
	BL-Jan	72.022	1.568
	BL-Feb	68.346	1.511
	BL-Mar	74.569	1.486
	PST-Sep	61.617	1.721
	PST-Oct	63.975	1.696
	PST-Nov	65.957	1.732
	PST-Dec	69.088	1.904
	PST-Jan	66.059	1.775
	PST-Feb	65.006	1.760
	PST-Mar	63.582	1.795

Descriptive Statistics 2x7 Baseline Only Analysis

In the baseline year, September through March, 2010 – 2011, there were a total of 593 (*N*=593) total cases possible for the study when students were included in the attendance of the 19 teachers instructing students in grades 2 through 5 in both the comparison and treatment groups.. Four cases (.67% of total cases) were removed from the original data set due to the students moving between schools causing their data to exist in both the comparison and treatment group's data. In addition, two cases (.34% of total cases) were removed due to the cases missing two or more scores out of the seven possible data points. Twenty-six cases (4.38%) were missing one score of the possible seven data points. Missing data was imputed by taking the mean of the non-missing values. One-hundred and eighty-six cases (29.67%) were removed due to the cases only presenting a post set of scores. One-hundred Eighty-two cases

were removed because these cases did not have post period data (e.g. because they were in the sixth grade during the post period, or transferred to another school, etc.) resulting in 411 valid cases (*n*=411) with 5754 data points for analysis of the pretest data or baseline year.

Table 5 provides the number of cases assessed each month, the range of scores attained; the mean score attained for each month, the standard deviation for each mean score, as-well-as the skewness and kurtosis figures. The descriptive data in table 5 represents combined data derived from both groups. The mean score of both groups increased from September through March (63.47 to 75.05) due to the test design to measure end of year learning. You would expect to see an increase in the scores over time due to the natural learning curve of the students. Given the number of cases occurring, the standard deviation is within expectations and there is no significance in the skewness or kurtosis data.

Table 6

Descriptive Statistics of baseline Scores for the 2010 – 2011 School Year

Assessment Month	n	Min	Max	Mean	Standard Deviation	Skewness		Kurtosis	
	Stat	Stat	Stat	Stat	Stat	Stat	Std Error	Stat	Std Error
September	411	10	100	63.47	18.538	-.453	.120	-.395	.240
October	411	15	100	71.35	18.767	-.454	.120	-.493	.240
November	411	10	100	66.16	20.289	-.665	.120	-.221	.240
December	411	15	100	75.70	18.499	-.939	.120	.343	.240
January	411	15	100	73.44	19.202	-.775	.120	-.111	.240
February	411	10	100	72.28	18.432	-.810	.120	.226	.240
March	411	13	100	75.05	17.990	-.969	.120	.768	.240

Additional descriptive statistical information is presented in the appendix section providing pretest and posttest histograms of the distribution of scores and frequency of combined group scores for each month (see appendix k on page 107).

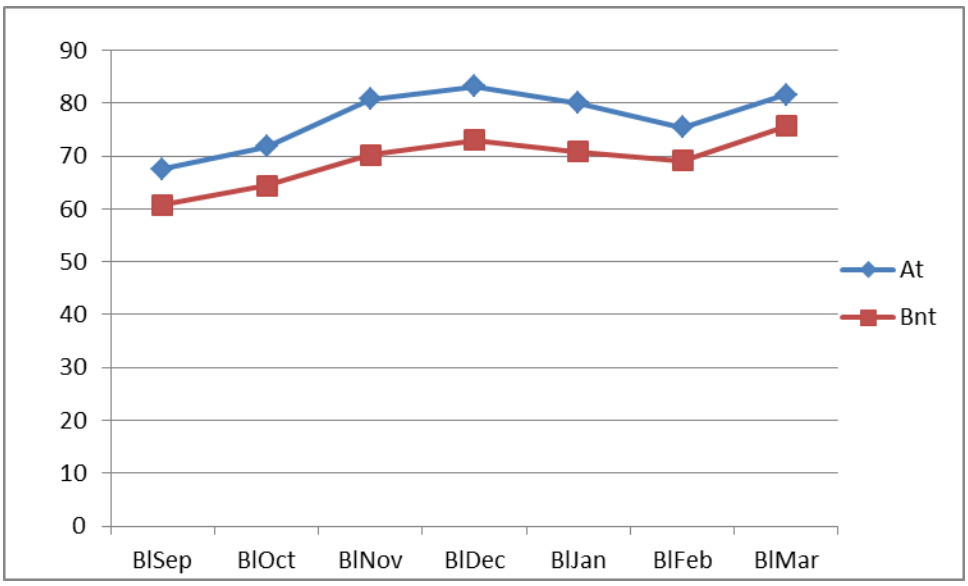
Data and Analysis ANOVA 2 x 7 Baseline Only

An analysis of variance between mean scores was completed on the baseline data from the 2010 – 2011 school year. This was conducted using a 2 x 7 ANOVA based upon two groups being tested monthly over a seven month period of time. The within-subjects factors included an independent variable of time and a dependent variable of monthly test scores. There were seven months of tests scores inclusive of the months September to March. The between-subjects factors were the cases associated with each school. The comparison building (B_{nt}) and the building in which the teachers participated in the yearlong professional development treatment (A_t). In B_{nt} n = 232 and in A_t n = 179.

In Graph 1 the relationship of the mean scores running from September to March indicates that group A_t realized a mean score higher than group B_{nt}. Through all seven data points the relationship between the mean scores for both groups remains consistent in that each data point rises and falls in unison, and the distance between each mean score for each data point does not vary by more than 1.53 points. The mean score for both buildings increase as would be expected. The increase in the mean scores reflects learning over time. The E-valuate tool is an end of the year test measuring the learning over time of end of year grade level expectations. What is significant in the data is both groups' achievement scores behaved in the same way. In the post test, the two groups' mean scores did not behave in the same way (See table 7 on page 127 for detailed scores for each month in the baseline year). Questions as to what affected the mean scores in group A_t and B_{nt} are of interest.

Graph 3

Baseline Mean Scores for Groups A_t and B_{nt} base year



A multivariate analysis of variance between the mean score of groups A_t and $B_{nt,}$ a Wilks' Lambda test, was utilized to evaluate the Time x Group interaction effect for the baseline year. Exact statistics were utilized, and the test was computed using an alpha of .05. The test revealed there was not a significant effect for Time x Group $\Lambda = .994$, $F\ (6.0, 404) = .378$, $p = .893$, $\eta^2 = .006$ and an observed power of .160. The test run for time interaction resulted in an effect. An effect was expected as learning over time would be expected. The E-valuate tool measures learning of GLEs over time. The test resulted in $\Lambda = .564$, $F\ (6.0, 404) = 52.101$, $p = .000$, $\eta^2 = .436$,-observed power of 1.000. The F_{crit} and p-values support the effect of learning over time. The Mauchly's Tests of Sphericity was run to test the repeated measures analysis of variance assumption of sphericity. This test was conducted to test for inflated p-values for the F statistics due in part to the error covariance matrix of the orthonomalized transformed variables being disproportional relative to an identity matrix (1s on the diagonal, Os elsewhere). The test resulted in $X^2(20) = 693.501$, $p = .000$, $\dot{\varepsilon} = .525$. The test indicates the assumption of sphericity

was violated (Mauchly's w = .182 and € = .525). Since Mauchly's Test of Sphericity is highly sensitive to even minor deviations from sphericity, p-values of less than .001 (as indicated here) are generally interpreted as significant violations of the assumptions. In these situations, adjustments to the degrees of freedom for the multivariate tests are made, and the Greenhouse-Geisser adjustment was made here to the Time x Group interaction effect. The test resulted in SS_x = 239.061, F (75.830, 3.153) = .197, p = .906, η^2 = .000, -observed power .088. With the adjusted F_{crit} and p-value statistics, the Greenhouse-Geisser confirms the lack of effect for time by group.

A test of within-subjects contrast (6 single degree-of –freedom polynomial contrasts represent the Time x Group interaction effect) was also completed. The Time x Group interaction was not significant for both linear trend (ESS = 22.137, F (1, 409) = .050, p = .824, η^2 = .000, observed power -.056), as evidenced in the B_{nt} and A_t trends in graph 3. The linear trend (ESS = 35111.726, F (1, 409) = .78.592, p = .000, η^2 = .161, observed power -1.000.) shows the effect over time was retained as evidenced in graph 3.

A test of between-subjects effects (ESS = 4592, 8.005, F (1, 409) = 39.465, p = .000, η^2 = .088, observed power -1.000. confirms the effect of learning over time remained.

An estimated marginal means was run to address the difference in sample sizes between the groups. This test also allowed for the adjustment of mean scores between groups collapsed across all seven data points. The estimated marginal means for group A_t resulted in M = 75616 (.964), 95%, CI (73.722, 77.511). For group B_{nt} M= 67.588 (.847), 95%, CI (65.894, 69.222). The estimate in marginal means test result quantifies the difference in mean scores between group A_t and B_{nt} over the seven data points. The analysis between group A_t and B_{nt} over the seven data points confirmed the effect of learning over time as expected, but no additional effect

was detected.

Descriptive Statistics 2x7 Post Only Analysis

In the post year, September through March, 2011-2012, the number of cases for the analysis was four-hundred seven, $N = 407$. Four cases (.84% of total cases) were removed due to the students moving between schools causing their data to exist in both the comparison and treatment group's data. Two cases (.33% of total cases) were removed due to the cases missing two or more scores out of the seven possible data points. Twenty-six cases (4.38%) were missing one score of the possible seven data points. Missing data was imputed by taking the mean of the non-missing values. One-hundred and seventy-six cases (29.67%) were removed due to the cases only presenting a post set of scores. The removal of these cases total 186 resulting in 407 valid cases (n=407) with 5698 data points for analysis of the post data or treatment year. Of the 407 cases in the treatment year, 164 (27.7%) cases were taught by teachers who participated in the yearlong professional development treatment. There were 243 (41.0%) cases taught by teachers who did not participate in the yearlong professional development treatment. There were 122 (20.6%) cases taught by teachers who utilized essential learning goals in their planning and instruction. Teaching experience ranged from 2.0 years to 27.0 years (**see** Table 8 on page 128). The number of years of education post bachelors degree ranged from 0.0 to 8.0 years (see Table 9 on page 129**).**

Table 6 provides the number of cases assessed each month, the range of scores attained; the mean score attained for each month, the standard deviation for each mean score, as well as the skewness and kurtosis figures. The descriptive data in table 6 represents combined data derived from both groups. The mean score of both groups increased from September through March (63.47 to 75.05) due to the test design to measure end of year learning. Again, you would

expect to see an increase in the scores over time due to the natural learning curve of the students. Given the number of cases occurring, the standard deviation is within expectations, and there is no significance in the skewness or kurtosis data.

Table 10

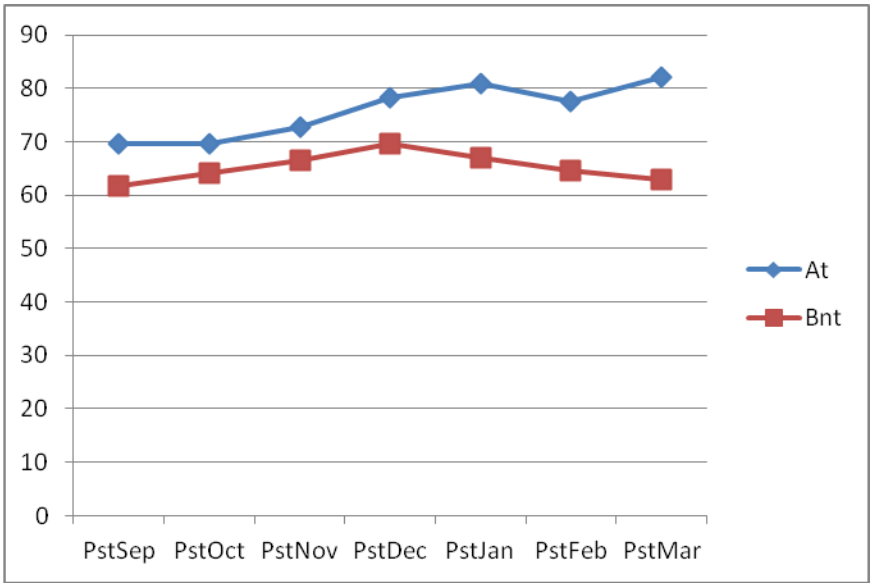
Descriptive Statistics of Posttest Scores for the 2011 – 2012 School Year

Assessment Month	n	Min	Max	Mean	Standard Deviation	Skewness		Kurtosis	
	Stat	Stat	Stat	Stat	Stat	Stat	Std Error	Stat	Std Error
September	407	10	100	64.26	18.538	-.453	.120	-.395	.240
October	407	15	100	65.64	18.767	-.454	.120	-.493	.240
November	407	15	100	69.09	20.289	-.665	.120	-.221	.240
December	407	10	100	74.37	18.499	-.939	.120	.343	.240
January	407	15	100	73.49	19.202	-.775	.120	-.111	.240
February	407	10	100	71.61	18.432	-.810	.120	.226	.240
March	407	10	100	73.44	17.990	-.969	.120	.768	.240

In Graph 2, the relationship of the mean scores running from September to March indicates group A_t realized a mean score increase greater than group B_{nt} over the seven data points from September through March. In September, group A_t scores resulted in M= 69.62, (18.778). In September, group B_{nt} scores resulted in M = 61.71, (19.953). The difference between the mean scores for groups $A_{\mathbf{t}}$ and B_{nt} in September of the post year, was 7.91 points. In March, group A_t scores resulted in M= 82.19, (14.432). In March, group B_{nt} scores resulted in M = 63.99, (23.245). The difference between the mean scores for groups $A_{\mathbf{t}}$ and B_{nt} in March of the post year, was 18.20 points (see Table 11 on page 131 for complete set of mean scores). Graph 2 represents the mean scores of group A_t and B_{nt} for the post treatment period over a seven month period of time. The difference in the mean scores for the 7 month period, from September to March, in the post year period, is clearly visible.

Graph 4

Baseline Mean Scores for Groups A_t and B_{nt} post year



Data and Analysis ANOVA 2 x 7 Post Only

To analyze for effect during the baseline year tests were run including the following: the multivariate Wilks' Lambda test, Mauchly's Test of Sphericity, the Greenhouse-Geisser within-subjects test, and a within-subjects contrast test. In addition, a transformed variable test was run, and an estimated marginal means test was run.

The Wilks' Lambda tests was utilized to identify multivariate differences in mean scores between the within groups. Exact statistics were utilized and was computed using an alpha of .05. The multivariate test resulted in $\Lambda = .808$, $F(6.0, 218) = .5.529$, $p = .000$, $\eta^2 = .132$, observed power - .996. The F_{crit} and p-values support the existence of significant effect associated with differential change in the study group over time (Furlong, Lovelace, & Lovelace, 2000). In addition, the Partial Eta Squared indicated significance with $\eta^2 = .132$ (Levine and Hullett, 2002).

Due to the sample sizes being different in size where A^t group $n = 85$ and B_{nt} group $n = 140$, a Box's M test of equality of covariance matrices was run to address departure from

multivariate normality given a p-value < .001. Box's M = 72.348, F (1, 28) = 2.537, p = 000. Box's M test, notoriously liberal, was used to test for significant problems with unequal variances in the multivariate ANOVAs. Given the unbalanced sample sizes and the obvious differences in variance (i.e., the scores of children in the A_t school were always higher and less variable than the scores of children in the B_{nt} school), Box's M, was significant in general. However, the p-value obtained for the Group x Time effect was p < .001, indicating a substantial effect despite any issues with multivariate normality assumptions. More importantly, the larger sample (B_{nt} cases) always had the larger variance, which means that any effect of multivariate nonnormality would serve to over-estimate the obtained p-values for effects, which means that p < .001 is a minimum estimate.

The Mauchly's Tests of Sphericity was run to test the repeated measure analysis of variance assumption of sphericity. This test was conducted to test for inflated p-values for the F statistics due in part to the error covariance matrix of the orthonomarlized transformed dependent variables being disproportional relative to an identity matrix (1s on the diagonal, 0s elsewhere). The Mauchlys' Tests of Sphericity resulted in $X^2(20) = 703$, p = .000, ἐ = .888. With the sensitivity of Mauchlys' Test of Sphericity to even minor deviation from sphericity, p-values of less than .001, (as indicated here) are generally interpreted as significant violation of the assumptions. Thus the test indicates that the assumption of sphericity was violated. An adjustment to the degrees of freedom for the multivariate test was made using the Greenhouse-Geisser to adjust the Time x Group interaction effect. This adjustment produces robust F and p – values despite the violation. The tests resulted in $SS_x = 6782.332$, F (1272.561, 5.330) = 8.07, p= .000, $\eta^2 = .035$, -observed power 1.00. Despite the violation of sphericity, the Time x Group interaction was retained as significant.

A test of within-subjects contrast (6 single degrees-of freedom polynomial contrasts represent the Time x Group interaction effect) was completed. The Time x Group interaction was significant (ESS = 5277.744, F (1, 223) = 24.555, p = .000, η^2 = .099, -observed power .999. The mean scores increased during the post year, September through March, in group $A_{t.}$ The within–subjects contrasts test supports the results of the Wilks' Lambda multivariate test.

An estimated marginal means was run to address the difference in sample sizes between the groups. This test also allowed for the adjustment of mean scores between groups collapsed across all seven data points. The estimated marginal means for group A_t resulted in M = 75.842 (.1.829), 95%, CI (72.219, 79.429). For group B_{nt} M = 65.337 (1.425), 95%, CI (62.528, 68.146). The estimate in marginal mean test results quantifies the difference in mean scores between group A_t and B_{nt} during the post year, September through March.

Summary

During the baseline year, both group A_t and B_{nt} mean scores ran approximately parallel to each other. Group A_t mean scores were higher for all seven data points than group B_{nt} mean scores for the baseline year. This split in values was constant across each of the seven data points. During the post year, the group A_t mean scores began to increase the split between the two groups starting in December and doubled in size by the end of the post year.

A 2 X 14 ANOVA analysis of variance was conducted with a variety of different tests between the two groups over time. The Wilks' Lambda test for the effect of time on learning (reading scores) resulted in λ = .383, F (13, 211) = 26.101, p = .000, η^2 = .617. More importantly, the Wilks' Lambda test for the Time x Group interaction resulted in λ = .850, F (13, 211) = 2.684, p = .001, η^2 = .150, which remained significant after adjustment to degrees of freedom. This effect indicated that in the post year, students in the At group steadily improved from September

to March, while students in the B_{nt} group began to improve early but then leveled off from December to March.

The tests run to address the explicit use of goals did not indicate an effect for that variable beyond the effect of learning over time. The covariates of years of experience and years of education could not be used to explain away the Time x Group interaction

The test run on the baseline year and post year separately in a pair of 2 x7 analysis of variance produced results consistent with the 2 x 14 analysis of variance. The F_{crit} and p-values for the baseline year did not reflect any effect. The F_{crit} and p-values for the post year indicated an effect had occurred for the treatment group A_t, in that those students continued to improve while the B_{nt} student performance leveled off.

Chapter V

Conclusions, Recommendations, Implications

Conclusions

The purpose of this study was to examine the effect of teacher professional development treatment on essential learning goals on grades two through five student reading achievement.

The results of the analysis of variance support the rejection of the null hypothesis. The F_{crit} and p-values in the tests run indicate that an interaction effect did take place. The covariate factors of goal use observed, teacher experience, and education beyond a bachelors' degree did not significantly impact the observed effect on student achievement scores.

Recommendations

Based upon the occurrence of an effect on the achievement scores in the area of reading for the students in grades two through five, it is the recommendation of the researcher additional research is conducted to confirm the effect of professional development for teachers in the area developing essential learning goals. There are also a number of limitations in the study to be addressed in further studies. Limitations such as difference in population size, and the social demographic differences between the two groups could be addressed. Additional areas in future studies should consider the possible effects of increased focus on assessment, use of assessment data to determine next learning steps, and the feedback process. Black and Wiliam (2010) have noted the positive effects on student achievement.

Another cause for further study is the fact that the B_{nt} group mean scores did not improve in the post year as shown in graph one on page 61. In the base year the B_{nt} experienced a spike in scores similar to group $A_{\mathbf{t}}$ from February to March; but in the post year, group B_{nt} failed to improve from February through March. One would question why the score patterns occurred as

they did. Would the A_t group mean scores have leveled off if their teachers had not participated in the professional development treatment?

It is the recommendations of the researcher that a larger group be utilized across a more diverse student demographic population. In addition, a longer period of time providing a professional development treatment that better meets teachers at a variety of readiness levels would be beneficial. Additional professional development time would allow for the professional development treatment to address teaching skills (development of improved assessments, analysis of assessment data, adjustments to lessons, determining next steps, and feedback to students) linked to the development of learning goals (Black and Wiliam, 2010; Wiggins and McTighe, 2008).

Implications

As school districts continue to be held accountable to improving school achievement scores, it will continue to be necessary for instruction to be more effective and efficient in improving student achievement. Strategies to improve instruction continue to be addressed at the government, university and local school district level. If further study could solidify the work of so many researchers that have worked in the area of developing our teachers knowledge of learning goals and their affected teaching strategies, policy at the national and district level could be impacted. If research could establish increasing the instructional capacity of our teachers in the areas of learning goals and associated teaching skills could significantly improve student learning, would a demand for a change in policy allow for the time necessary to improve our teachers' skills? Could that research result in a shift away from simply spending more time in the classroom to spending (supporting) more time to improve the instructional capacity or our teachers

References

Appalachia Educational Lab., C. (2005). *Module 7–Purposeful Student Assessment. School Improvement Specialist Training Materials: Performance Standards, Improving Schools, and Literature Review*. Appalachia Educational Laboratory at Edvantia. Retrieved from ERIC database.

Archived Information Document at Ed Gov V. Principals of high quality education, *Department of Education* 2000 http://www2.ed.gov/G2K/bridge.html

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85(6), 536-553.* BellSciEd01-85-536

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the Black Box: Assessment for Learning in the Classroom. *Phi Delta Kappan*, 86(1), 8. Retrieved from ERIC database.

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan, 80*(2), 139-44. Retrieved from ERIC database.

Black, P., & Wiliam, D. (2003). In Praise of Educational Research: Formative Assessment. *British Educational Research Journal*, 29(5), 623-637. Retrieved from EBSCO*host.*

Black, P., & Wiliam, D. (2010). "Kappan Classic": Inside the Black Box–Raising Standards through Classroom Assessment. *Phi Delta Kappan*, 92(1), 81-90. Retrieved from EBSCO*host*.

Bloom, B. S., & California Univ., L. n. (1968). *Toward A Theory of Testing Which Includes Measurement-Evaluation-Assessment. Center for the Study of Evaluation of Instructional Programs Occasional Report No. Nine*. Retrieved from EBSCO*host*.

Bloom, B., Chicago Univ., I., & Regional Educational Laboratory for the Carolinas and Virginia, D. (1968). *Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1*. Evaluation Comment. Retrieved from ERIC database

Bloom, B. S., & United Nations Educational, S. g. (France). International Inst. For Educational Planning. (1975). *Evaluation, Instruction and Policy Making. IIEP Seminar Paper: 9*. Retrieved from EBSCO*host*.

Bennett, K., & Cunningham, A. (2009). Teaching Formative Assessment Strategies to Preservice Teachers: Exploring the Use of Handheld Computing to Facilitate the Action Research Process. *Journal of Computing in Teacher Education*, *25*(3), 99-105. Retrieved from ERIC database.

Borek, J. (2008). "A Nation at Risk" at 25. *Phi Delta Kappan*, *89*(8), 572-574. Retrieved from ERIC database.

Borko, H. (1997). New Forms of Classroom Assessment: Implications for Staff Development. *Theory into Practice*, *36*(4), 231-38. Retrieved from ERIC database.

Borko, H. & Putnam, R. (1996). Learning to teach. In D. Berliner & R. Calfee (Eds.) *Handbook of educational psychology* (673-708). New York: MacMillan.

Cobb, P., McClain, K., Lamberg, T. & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher, 32*(6), 13-24.

Corcoran, T., Mosher, F. A., Rogat, A., & Consortium for Policy Research in, E. (2009). Learning Progressions in Science: An Evidence-Based Approach to Reform. CPRE Research Report # RR-63. *Consortium for Policy Research in Education*, Retrieved from

EBSCO*host*.

Consortium for Policy Research in, E. (2009). Taking Human Capital Seriously: Talented Teachers in Every Classroom, Talented Principals in Every School. Principles and Recommendations for the Strategic Management of Human Capital in Public Education. *Consortium for Policy Research in Education*, Retrieved from ERIC database.

DeCaro, S. A., (2003). Descriptive vs. Inferential statistics: What is the difference? Retrieved [January 9, 2011] from http://www.mymarketresearchmethods.com/descriptive-inferential-statistics-difference/

Darling-Hammond, L., & Baratz-Snowden, J. (2007). A Good Teacher in Every Classroom: Preparing the Highly Qualified Teachers Our Children Deserve. *Educational Horizons, 85*(2), 111-132. Retrieved from ERIC database.

Darling-Hammond, L., Cobb, V., Asia Pacific Economic Cooperation Secretariat & Department of Education, W. (1995). *Teacher Preparation and Professional Development in APEC Members: A Comparative Study*. Retrieved from ERIC database.

Department of Education, (2008). *Fiscal Year 2009 Performance Budget: Performance Plan*. US Department of Education. Retrieved from ERIC database.

Department of Education (ED), O. (2009). Great Expectations: Holding Ourselves and Our Schools Accountable for Results. *US Department of Education*, Retrieved from ERIC database.

DuFour, R. (2011, May). *Plcs getting started*. Presentaion presented at Professional learning communities: institute, Saint Charles, MO.

DuFour, R. DuFour R. Eacker, M., & Many, T. (2006). *Learning by doing: A handbook for professional learning communities as work*. Bloomington, IN: Solution Tree Press.

DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement.* Bloomington: Solution Tree Press.

DuFour, R., Eaker, R., & DuFour, R. (Eds.). (2005). *On common ground: The power of professional learning communities.* Bloomington Indiana: National Educational Services.

Dunn, K., & Mulvenon, S. (2009). Let's Talk Formative Assessment ... and Evaluation. *Online Submission*, Retrieved from ERIC database

Eaker, R., DuFour, R., & DuFour, R. (2002). G*etting started: Re-culturing schools to become professional learning communities.* Bloomington, Indiana: National Educational Service.

Everitt B.S. & Dunn G. (1991) *Applied Multivariate Data Analysis*. Edward Arnold. London. pp. 219-220.

Ferlong, N., Lovelace, E., & Lovelace, K. (2000). *Research methods and statistics: An integrated approach*. Belmont, CA: Wadsworth/Thomson Learning

Ginsburg, A., Leinwand, S., Anstrom, T., Pollock, E., & American Institutes for Research, W. (2005). What the United States Can Learn from Singapore's World-Class Mathematics System (and What Singapore Can Learn from the United States): An Exploratory Study. *American Institutes for Research*, Retrieved from ERIC database.

Guskey, T R (Sept 1994). Defining the Differences Between Outcome-based education and mastery learning. *School Administrator*, 51, 8. P.34. Retrieved, August 9, 2009, from Educator's Reference Complete via Gale: http://find.galegroup.com/gtx/start.do?prodId=PROF&userGroupName=morenetsccc

Heritage, M., Kim, J., Vendlinski, T., Herman, J., & National Center for Research on Evaluation, S. (2008). From Evidence to Action: A Seamless Process in Formative Assessment? CRESST Report 741. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST),* Retrieved from ERIC database.

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., Timms, M., & National Center for Research on Evaluation, S. (2011). Relationships between Teacher Knowledge, Assessment Practice, and Learning–Chicken, Egg, or Omelet? CRESST Report 809. *National Center For Research On Evaluation, Standards, And Student Testing (CRESST),*

Hess, K. (2007). Developing and Using Learning Progressions as a Schema for Measuring Progress. *National Center for Assessment/NCIEA*, Dover NH. http://www.nciea.org/publications/CCSSO2_KH08.pdf

Ketterlin-Geller, L. (2005). Knowing What All Students Know: Procedures for Developing Universal Design for Assessment. *Journal of Technology, Learning, and Assessment*, *4*(2), Retrieved from ERIC database.

Kohn, A. (1999, March). From degrading to de-grading. High School Magazine, 38-43.

Levine, T.R. & Hullett, C.R. (2002). Eta squared, partial eta squared and misreporting of effect size in communication of research. Human Communication Research, (28), p. 612-625.

Lewis, A., & California Univ., L. (2006). Celebrating 20 Years of Research on Educational Assessment: Proceedings of the 2005 CRESST Conference. CSE Technical Report 698. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST),* Retrieved from ERIC database.

Marzano, R. L. (2000). *Transforming classroom grading*. Alexandria, VA: Association for

Supervision and Curriculum Development.

Marzano, R. (2009). *Designing & teaching & learning goals & objectives: Classroom strategies that work*. Bloomington, IN: Marzano Research Laboratory.

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

McCowan, R., Center for Development of Human, S., & State University of New York (SUNY), R. (1998). Origins of Competency-Based Training. *Online Submission*, Retrieved from ERIC database.

Missouri Outstanding Education Act SB287, (1993). (http://openstates.org/mo/bills/2013/SB287/documents/MOD00006993/)

Missouri Department of Elementary and Secondary Education, (2006). Revised State Plan Highly Qualified Teachers, (www2.ed.gov/nclb/landing.jhtml).

Missouri Department of Elementary and Secondary Education, (2006). State Plan: High Quality Professional Development, (http://dese.mo.gov/divimprove/pd/hqpd.pdf)

Missouri Department of Elementary and Secondary Education, (2011) Missouri comprehensive data systems. (http://mcds.dese.mo.gov/quickfacts/SitePages/DistrictInfo).

Missouri Department of Elementary and Secondary Education, (2013) Teacher Evaluation: Missouri's Educator Evaluation System. (http://dese.mo.gov/eq/documents/eq-ees-teacher-evaluation.pdf), Retrieved from google search.

Missouri Department of Elementary Education and Secondary Education, (2013). Model Teacher and Leader Standards, (http://dese.mo.gov/eq/documents/TeacherStandards.pdf)

Mullis, I., Martin, M., Gonzalez, E., Chrostowski, S., International Association for the Evaluation of Educational, A., & Boston Coll., C. (2004). *TIMSS 2003 International Mathematics*

Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. TIMSS & PIRLS International Study Center. Retrieved from ERIC database.

National Association of Elementary School Principals, A., & National Staff Development Council, O. (1995). *Standards for Staff Development: Elementary School Edition. Study Guide*. Retrieved from ERIC database

Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company.* New York: Oxford Press.

National Commission on Teaching & America's Future, N. (1996). *What Matters Most: Teaching for America's Future. Report of the National Commission on Teaching & America's Future*. Retrieved from ERIC database.

National Council for Accreditation of Teacher Education (NCATE). (2002). *Professional standards for the accreditation of schools, colleges, and departments of education*. Washington, DC: Author.

National Council on Measurement in Education, W., American Federation of Teachers, W., & National Education Association, W. (1990). *Standards for Teacher Competence in Educational Assessment of Students*. Retrieved from ERIC database.

National Staff Development Council, (2001). *Standards for staff development, revised*. Retrieved November 30, 2011, from http://www.nsdc.org

Nicol, D., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education*, *31*(2), 199-218. Retrieved from ERIC database.

Petrilli, M. J., Scull, J., & Thomas B. Fordham, I. (2011). American Achievement in International

Perspective. *Thomas B. Fordham Institute*,

Popham, W. (2006). Assessment for Learning: An Endangered Species. *Educational Leadership, 63*(5), 82-83. Retrieved from ERIC database.

Pryor, J., & Crossouard, B. (2008). A Socio-Cultural Theorization of Formative Assessment. *Oxford Review of Education, 34*(1), 1-20. Retrieved from ERIC database

Rahn, M., & Stecher, B. (1997). Making decisions on assessment methods: Weighing the tradeoffs. *Preventing School Failure*, *41*(2), 85. Retrieved from Academic Search Premier database.

Richards, M. (2009). Closing the Gap between Professional Development and Student Achievement. *Online Submission*, Retrieved from ERIC database.

Rothman, R., & National Education Goals Panel, W. (2000). Bringing All Students to High Standards: Report on National Education Goals Panel Field Hearings. Lessons from the States. Retrieved from ERIC database.

Rudner, L. & Schafer, W. (2002). *What teachers need to know about assessment.* Washington DC: National Education Association.

Sadler, R**.** (1989). Formative assessment and the design of instructional systems. *Instructional Science 18(1), 119-144.* www.springerlink.com/index/ X7L185036H762M45.pdf

Schleicher, A., & Stewart, V. (2008). Learning from World-Class Schools. *Educational Leadership*, *66*(2), 44-51. Retrieved from ERIC database.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review, 57 (1), 1-22. Retrieved from Google search.

Stiggins, R. (2002). Assessment Crisis: The Absence of Assessment for Learning. *Phi Delta Kappan, 83*(10), 758-65. Retrieved from ERIC database.

Stiggins, R.J., Arter, J., Chappuis, J., & Chappuis, S. (2005). *Classroom Assessment for Student Learning: Doing It Right—Using It Well.* Portland OR: Assessment Training Institute Inc..

Sweeney, D. (2007). Mirror, Mirror, in the Lab: Process Shows Coaching Clear Reflections of Their Own Practices. *Journal of Staff Development, 28(1), 38-41*

Thompson, M., Goe, L., & Educational Testing, S. (2009). Models for Effective and Scalable Teacher Professional Development. Research Report. ETS RR-09-07. *Educational Testing Service,* Retrieved from ERIC database

United States Department of Education, (2009). No Child Left Behind-*Ed. Gov*. (www2.ed.gov/nclb/landing.jhtml).

Warren, J. & Grodsky, E. (2009). Exit Exams Harm Students Who Fail Them and Don't Benefit Students Who Pass Them. *Phi Delta Kappan, 90*(9), 645-649. Retrieved from ERIC database.

Wenglinsky, H., & Educational Testing Service, P. r. (2000). How Teaching Matters: Bringing the Classroom Back into Discussions of Teacher Quality.

Wiggins, G., & McTighe, J. (2011). *The understanding by design guide to creating high-quality units*. Alexandria, VA: ASCD

Wiggins G, McTighe J. Put Understanding First. *Educational Leadership* [serial online]. May 1, 2008; 65(8):36-41. Available from: ERIC, Ipswich, MA. Accessed September 13, 2011.

Wylie, E., Lyon, C., Goe, L., & Educational Testing, S. (2009). Teacher Professional Development Focused on Formative Assessment: Changing Teachers, Changing Schools.

Research Report. ETS RR-09-10. *Educational Testing Service*, Retrieved from ERIC database.

Appendix A

Instructional Competencies Identified

1. Teachers should possess a strong knowledge background and understanding of the curriculum they are going to teach
2. Teachers should possess skills essential in determining essential learning outcomes.
3. Teacher should possess knowledge of various effective pedagogy
4. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions**.**
5. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
6. The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.
7. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
8. Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.
9. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
10. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Appendix B

Three Essential Components
High Quality Professional Development

Part I definitions:

1. Actively engages teachers in planning, skills, and information over time. (Standard 2)
2. Is directly linked to improved student learning so that all children may meet the Show-Me Standards at the proficient level. (Standards 8, 10)
3. Is directly linked to district and building school improvement plans. (Standard 1)
4. Is developed with extensive participation of teachers, parents, principals, and other administrators [Parent participation maybe at the Comprehensive School Improvement Plan (CSIP) level]. (Standards 1, 2, 9, 12)
5. Provides time and other resources for learning, practice, and follow-up. (Standards 3, 7)
6. Is supported by district and building leadership. (Standard 2)
7. Provides teachers with the opportunity to give the district feedback on the effectiveness of participation in this professional development activity. (Standard 5) (p. 2)

Part II approved activities:

1. Study groups (Standard 1)
2. Grade-level collaboration and work (Standards 1, 9)
3. Content-area collaboration and work (Standards 1, 9)

4. Specialization-area collaboration and work (Standards 1, 9)
5. Action research and sharing of findings (Standards 4, 6)
6. Modeling (Standards 8, 9)
7. Peer coaching (Standards 8, 9)
8. Vertical teaming (Standards 1, 9)
9. Other (p. 2)

Part III identifies topics that include, but are not limited to, the following:

1. knowledge related to standards and classroom instruction, (Standard 11)
2. instructional strategies related to the content being taught in the classroom, (Standard 7)
3. improving classroom management skills, (Standards 9, 10)
4. content knowledge and content-specific teaching skills, (Standards 7, 11)
5. the integration of academic and career education, (Standard 9)
6. research-based instructional strategies, (Standards 6, 11)
7. strategies to assist teachers in providing instruction to children with limited English proficiency to improve their language and academic skills, (Standard 10)
8. strategies to assist teachers in creating and using classroom assessments, (Standard 5)
9. instruction in the use of data to inform classroom practice, (Standards 4, 11)
10. instruction in methods of teaching children with special needs, (Standard 10)
11. instruction in linking secondary and post-secondary education, (Standard 9)
12. involving families and other stakeholders in improving the learning of all students, (Standards 10, 12)

13. strategies for integrating technology into instruction, (Standard 10)
14. research and strategies for the education and care of preschool children, (Standard 6)
15. research and strategies for closing achievement gaps between diverse groups of students, (Standard 10) and
16. Other

Appendix C

Rudner and Schafer (2002) Recommendations

Necessary Knowledge and Understanding Necessary for the Development of Effective and Efficient Assessment

1. Administrators and teachers must understand the importance their own judgments in assessment areas such as "construction of test questions, scoring essays, creating rubrics, grading, or interpretation of results" (p. 6).
2. Administrators and teachers must understand the essential language of assessment measurement such as "the ability to understand and interpret the meaning of descriptive statistical procedures, including variability, correlation, percentiles, standard scores, growth-scale scores, norming, and principles of combining scores for grading" (p. 7). This knowledge is necessary for educators to communicate concerning assessment results.
3. Administrators and teachers must understand the purpose for the assessment. This is necessary to assure assessment aligns with the goals of the learning unit or course.
4. Administrators and teachers must understand the effects differing types of assessments have on student motivation and learning. Students are more engaged with some types of assessments, such as open response, and will study differently based upon the type of assessment method. It must be recognized studying differently may impact the way the material is learned.
5. Administrators and teachers must understand error occurs in all types of assessment. For this reason educators must understand how reliability is determined and how much error can occur in their assessments. It is also "critical that all educators

understand concepts like standard error of measurement, reliability coefficients, confidence intervals, and standard setting" (p. 8).

6. Administrators and teachers must understand assessments must be integrated with instruction. Assessment can no longer be utilized to simply audit learning.
7. Administrators and teachers must understand the concept of validity. Understanding validity is important in assisting educators in "making reasonable and appropriate inferences" concerning assessment results (p. 9).
9. Administrators and teachers must utilized assessment in a fair and ethical manner. Assessment must be absent of bias, equitable in nature, possess equality in outcomes, and utilized as an opportunity to learn.
10. Administrators and teachers must keep assessment "efficient and feasible" to assure utilization in the instruction process (p. 10).
11. Administrators and teachers must be proficient in the utilization of technology in implementation of assessment in the classroom. Utilizing technology will promote student engagement in the assessment process and can be a useful tool to promote efficiency in the assessment process.

Appendix D
Reliability, Validity, & Difficulty:
Benchmark Assessment Content Validation Process
E-valuate Assessment Tool

The technology platform of an assessment system, the user interface of that system, the amount of bells and whistles that system has – these aspects of an assessment system ensure that a support tool will provide users a non-frustrating experience. But these aspects of the system will not guarantee success; only the strength of an assessment system's content will determine that. Of utmost importance are three aspects of the content: its reliability, its validity, and its difficulty.

Reliability

Reliability is a measure of how consistently difficult questions are each month, and within individual standards and skills. To determine the reliability of each item in our item bank, we will use the following process:

1. Each item will be tagged to identify what state it is, what standard is being covered, what specific objective is being measured, what month the question is scheduled to be used in, what subject is being tested, and what grade level the question is used in.

2. Questions will then be sorted by state and grade and subject. This creates item pools of 180-225 questions.

3. Inter-item correlations will then be run. Inter-item correlations are measured by using what is called a Cronbach-alpha analysis (or Kuder-Richardson, to be more specific). This analysis produces a statistical measure, an alpha, for each individual question. The analysis also creates a matrix that identifies which items (when removed) would increase the item set's overall alpha. The sample student population sample for this analysis will be between 100-150 randomly-

selected students (per state, subject, and grade).

4. Items whose alpha are below 0.5000 (on a scale of 0.0000 to 1.0000) will be flagged for the content team to examine. Items whose removal would raise the overall item pool alpha by more than two standard deviations also will be flagged. The content team will have final determination as to whether or not a flagged item directly mirrors a particular state assessment item enough to keep, revise, or replace said item. Ideally, overall item pool alphas will not fall below 0.6500

Validity

Validity is a measure of how well performance on benchmark assessments matches up against performance on high-stakes assessments. To determine the validity of each item in our item bank, we will use the following process:

1. Individual student performance on high-stakes assessments is collected for each available state in the Edison system. The data from these assessments that will be used in correlations is the individual students' scale scores (by which states set thresholds of performance). The student population sample for this analysis is determined by the number of Edison schools within the state (all students will be used).

2. Individual student benchmark performance will be exported from individual school servers. The data to be used in the correlation is the overall student score on a particular benchmark. The benchmarks to be used are the most recent two benchmarks prior to the month in which the state assessment is administered. For example, if a state assessment is administered in April, benchmark scores from March and April (or February and March if students did not take April benchmarks due to testing) will be used. Their two months' worth of scores will be averaged together.

3. A simple regression analysis will be run matching individual student performance on high-stakes assessments to their average individual performance on benchmarks. The equation to the "best fit" line of the regression will allow us to determine the benchmark thresholds which will guarantee proficiency on state assessments to two standard deviations (95% confidence). Correlations will be run in every state, by subject and by grade. Generally, the r-squared of our correlations fall between 0.500-0.750.

Difficulty

Difficulty is a measure of how well students perform on a particular item compared to how well those same students perform on another item (regardless of standard). Each item in our item bank will be assigned a number from 0.000 to 1.000 that signifies how difficult the question is – the higher the figure, the more difficult the item. This Difficulty Index (DI) will be calculated using the following formulae:

1. We will run queries using the Data Warehouse (or receive an extract from the Skokie Team) that will provide us what percentage of students across a randomly-selected sample of students (per state, per subject, per grade) correctly answered each individual item. This overall percentage correct will constitute potentially one-quarter of the total DI according to the following scale, where DI_C represents the Difficulty Index Correct value (each item will receive a DI_C figure): Percentage of students correctly answering itemDI_C25% or less0.2525% - 40%0.2040% - 50%0.1550% - 60%0.1060% - 75%0.0575% or above0.00

2. Using the same student sample performance, we will sort student performance based on the student's overall percentage correct on that month's assessment. We will then pull out the performance of all students with total scores of 75% or higher on an entire month's assessment. We will then examine what percentage of this higher-performing group of students correctly

answered an individual item. This overall percentage correct in the top quartile will constitute potentially thirty-five percent of the total DI according to the following formula:

$$DI_{TQ} = (125\% - Z_{TQ}\%) \times 0.35$$

where DI_{TQ} represents the Difficulty Index Top Quartile value and Z_{TQ} represents the percentage of top quartile students correctly answering that particular item. Each item will receive a DI_{TQ} figure. We will correct for improved learning as the year progresses by averaging the progress made over the year and to achieve a Learning Coefficient (LC) and subtracting LC from Z_{TQ}. LC will be determined by averaging system wide performance in September-October averages from April-May averages and dividing by 9.

3. Using the same student sample performance, we will sort student performance based on the student's overall percentage correct on that month's assessment. We will then pull out the performance of all students with total scores of 25% or lower on an entire month's assessment. We will then examine what percentage of this lower-performing group of students correctly answered an individual item. This overall percentage correct in the Bottom quartile will constitute potentially twenty-five percent of the total DI according to the following formula:

$$DI_{BQ} = (125\% - Z_{BQ}\%) \times 0.25$$

where DI_{BQ} represents the Difficulty Index Bottom Quartile value and Z_{BQ} represents the percentage of bottom quartile students correctly answering that particular item. Each item will receive a DI_{BQ} figure. We will correct for improved learning as the year progresses by averaging the progress made over the year and to achieve a Learning Coefficient (LC) and subtracting LC from Z_{BQ}. LC will be determined as above.

4. Using the same student sample performance, we will sort student performance based on the student's overall percentage correct on that month's assessment. We will then pull out the

performance of all students with total scores of 25% - 75% on an entire month's assessment. We will then examine what percentage of this middle-performing group of students correctly answered an individual item. This overall percentage correct in the Middle quartiles will constitute potentially fifteen percent of the total DI according to the following formula:

$$DI_{MQ} = (125\% - Z_{MQ}\%) \times 0.15$$

where DI_{MQ} represents the Difficulty Index Middle Quartiles value and Z_{MQ} represents the percentage of middle quartiles students correctly answering that particular item. Each item will receive a DI_{MQ} figure. We will correct for improved learning as the year progresses by averaging the progress made over the year and to achieve a Learning Coefficient (LC) and subtracting LC from Z_{MQ}. LC will be determined as above.

5. Each item will then receive an overall Difficulty Index (DI) that is the sum of the above four measures, or

$$DI = DI_C + DI_{TQ} + DI_{BQ} + DI_{MQ}$$

6. The content team will receive the DI for each individual item by which they can calculate the DI of entire tests by adding up the individual DI of each item in that month's assessment, allowing them to better balance difficulty between months.

Appendix E

Instrument Utilized for Observations of Teacher Classroom Instruction

Teacher T1….	Yes	No
Learning goal posted	X	
Goal addressed in plan book	X	
Goal addressed during lesson	X	
Observation value assigned	1	

Appendix F

Form Utilized to Track Individual Classroom Observation Values Assigned from Observer

Teacher	Obs 1 Value	Obs 2 Value	Obs 3 Value	Obs 4 Value	Obs 5 Value	Factor Value Assigned
T1	**1**	**1**	**0**	**1**	**0**	**1**
T2	**1**	**1**	**1**	**1**	**1**	**1**
T3	**0**	**1**	**0**	**1**	**0**	**0**
T4	**0**	**0**	**0**	**0**	**0**	**0**
T5	**1**	**0**	**0**	**0**	**1**	**0**

Appendix G

Template Utilized to Assist Teachers in Aligning Learning Progressions to Topics and Skills

Literal Comprehensions	Prediction	Context Clues	Connections to… Text, Self, World	Retell and Summary	Beginning, Middle, End Summary	Main Idea and Supporting Details
LP Step 1						
LP Step 2						
LP Step 3						

Appendix H

Learning Lab Observation Sheet

Teacher Requesting Learning Lab: _____

Date of learning lab: _____

Time of learning lab: _____

Teachers observing: _____

_____

Instructional practice to review	Teachers responsibility	Observers responsibilities
What Went Well	I have a question(s)	Have you thought of
Teacher next steps	Group next steps	

Appendix I

Ambrose (1987) Complex Change Model

Vision	Skill	Incentive	Resource	Action Plan	=Change
	Skill	Incentive	Resource	Action Plan	=Confusion
Vision		Incentive	Resource	Action Plan	=Anxiety
Vision	Skill		Resource	Action Plan	=Gradual Change
Vision	Skill	Incentive		Action Plan	=Frustration
Vision	Skill	Incentive	Resource		=False Start

Appendix J

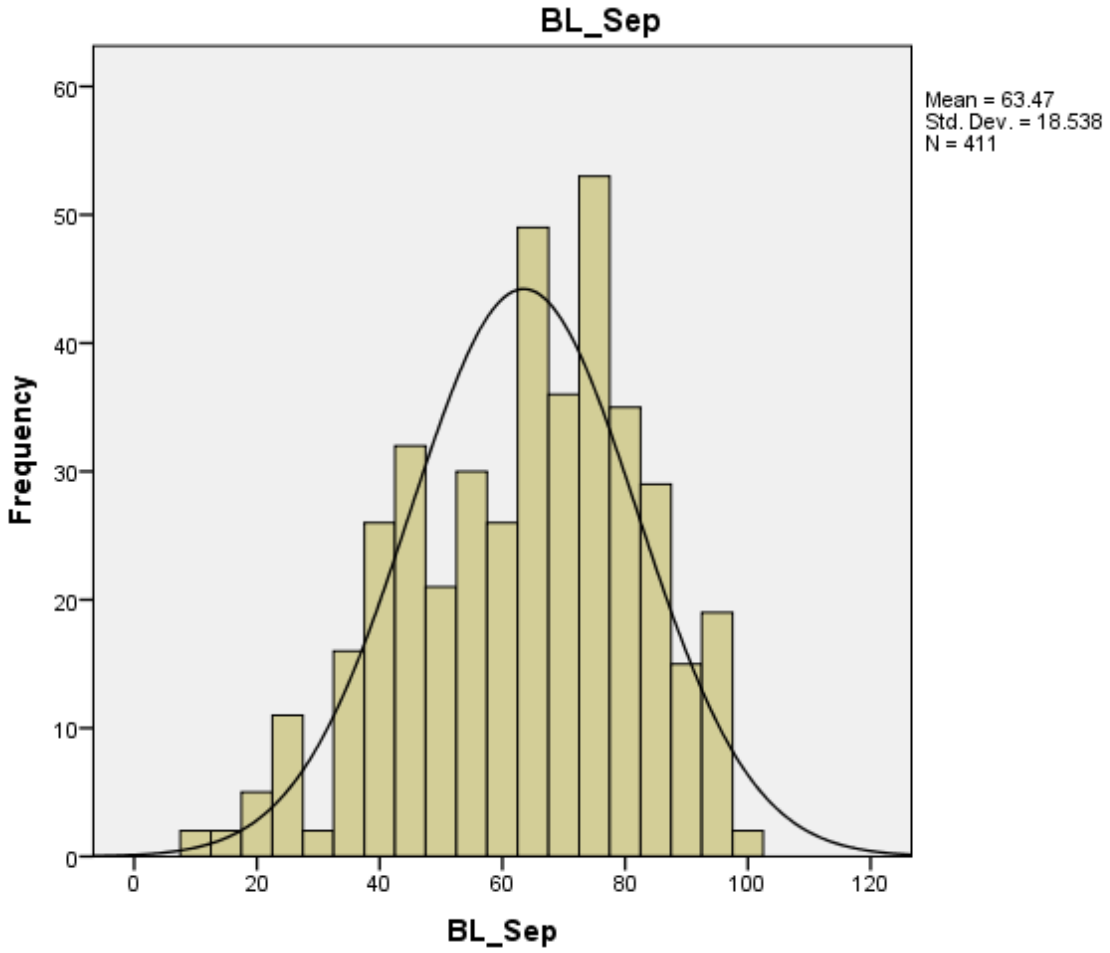
Observation of Goal Use

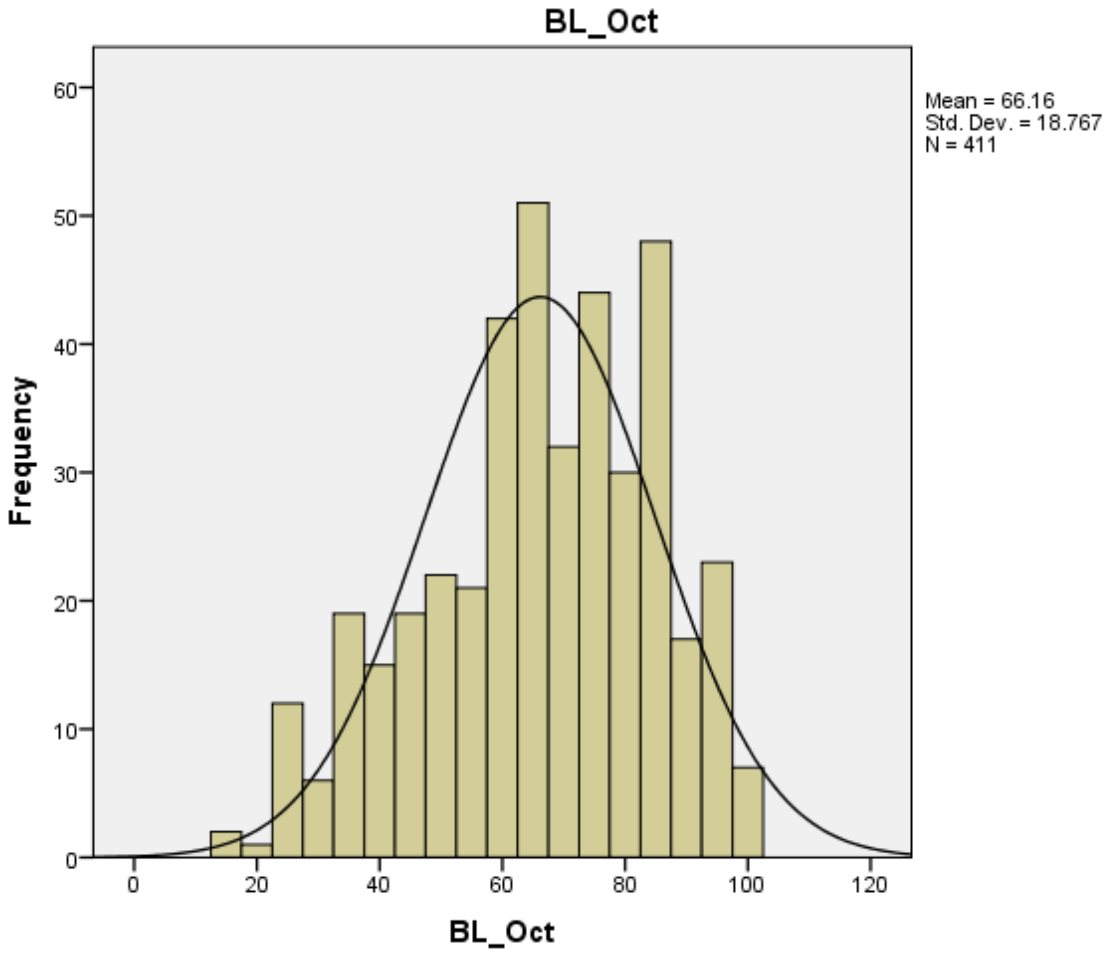
Values Assigned

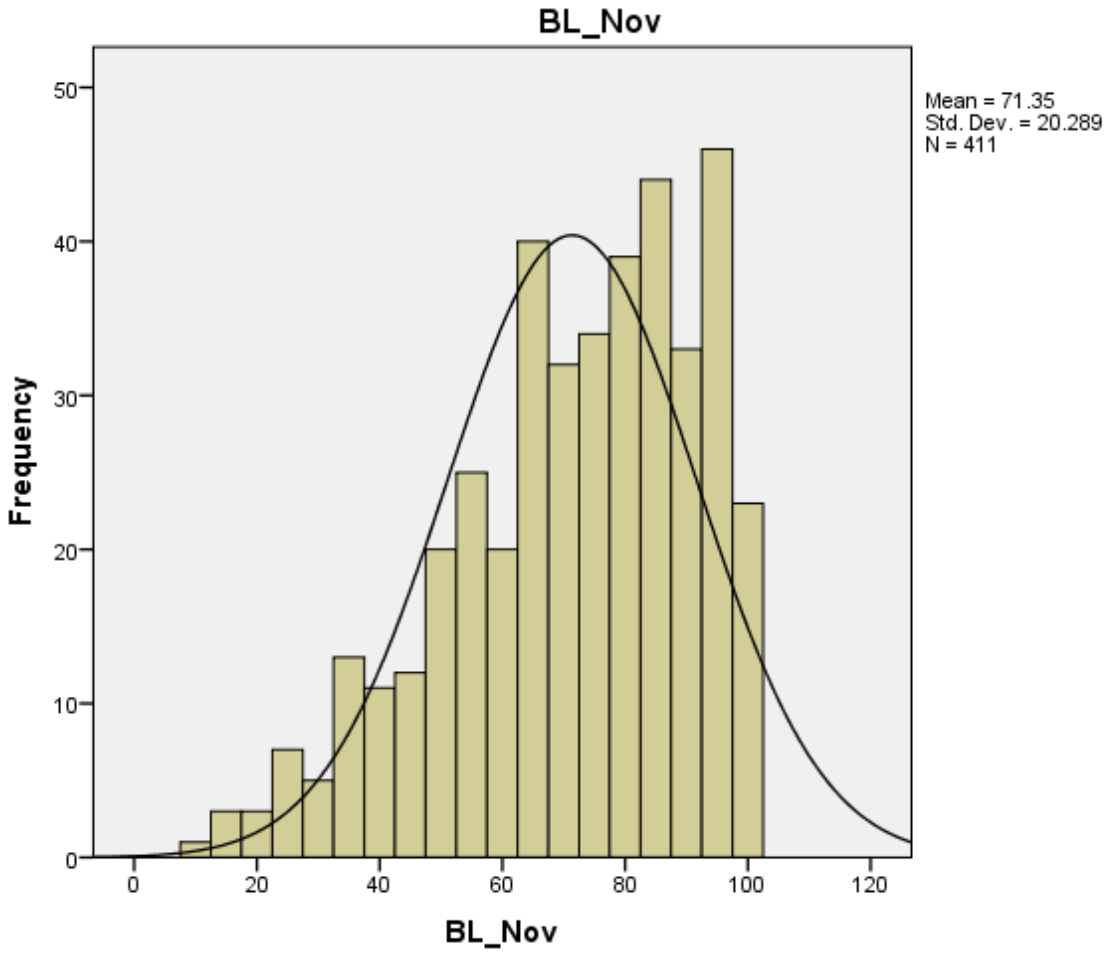
Teacher	Obs 1 Value	Obs 2 Value	Obs 3 Value	Obs 4 Value	Obs 5 Value	Factor Value Assigned
T1	**1**	**1**	**0**	**1**	**1**	**1**
T2	**1**	**1**	**1**	**1**	**1**	**1**
T3	**0**	**1**	**1**	**1**	**1**	**1**
T4	**1**	**1**	**1**	**1**	**1**	**1**
T5	**1**	**1**	**1**	**1**	**1**	**1**
T6	**1**	**0**	**1**	**1**	**1**	**1**
T7	**0**	**0**	**1**	**0**	**1**	**0**
T8	**1**	**1**	**1**	**1**	**1**	**1**
T9	**1**	**0**	**1**	**1**	**0**	**1**

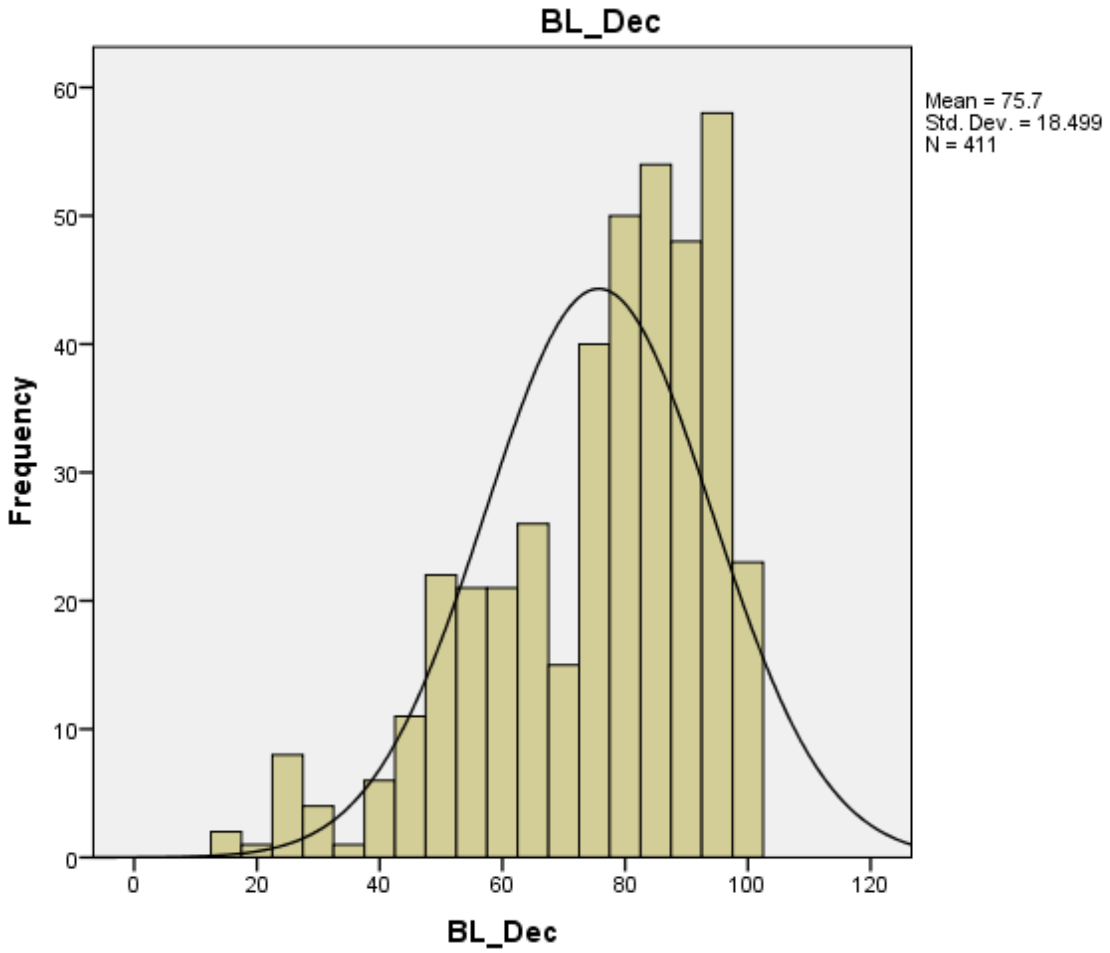
Appendix K

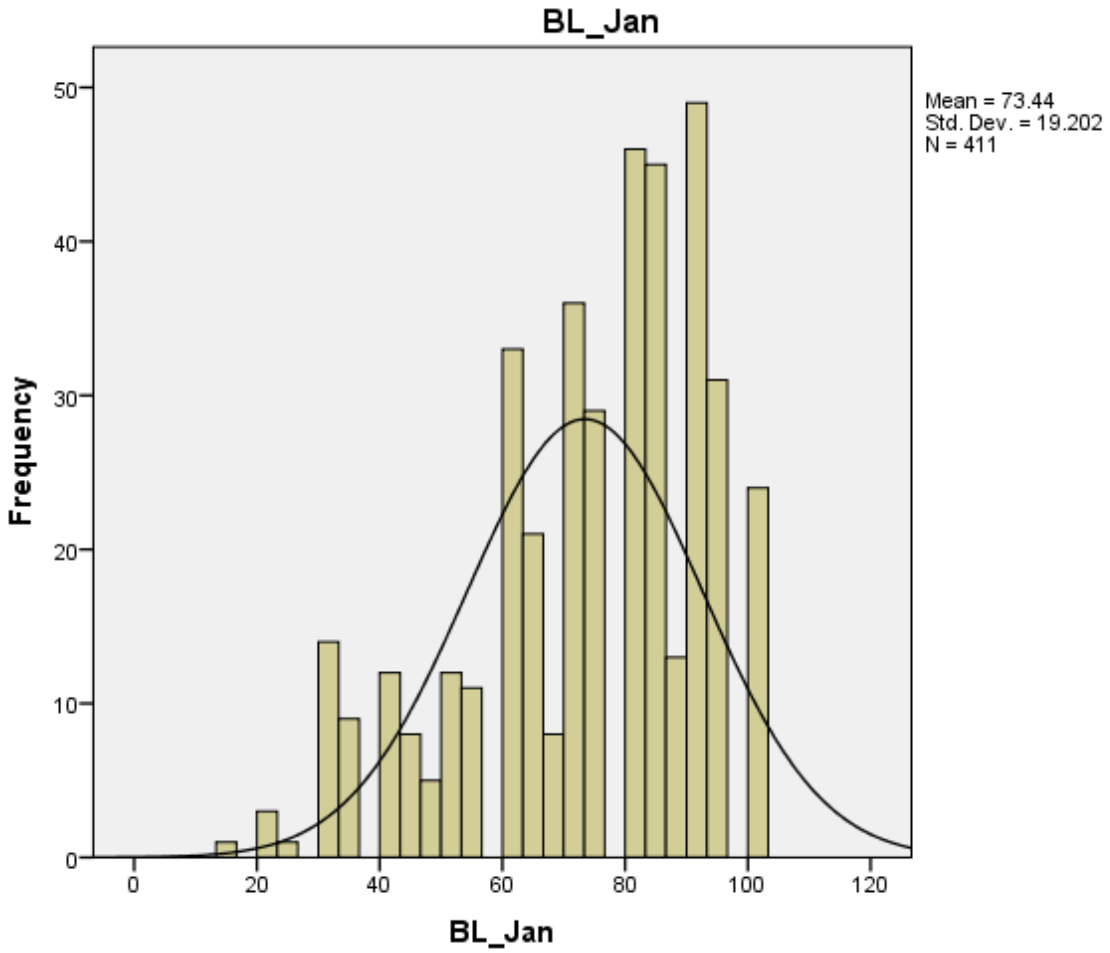
Histograms of Monthly Scores Both Baseline and Post Years

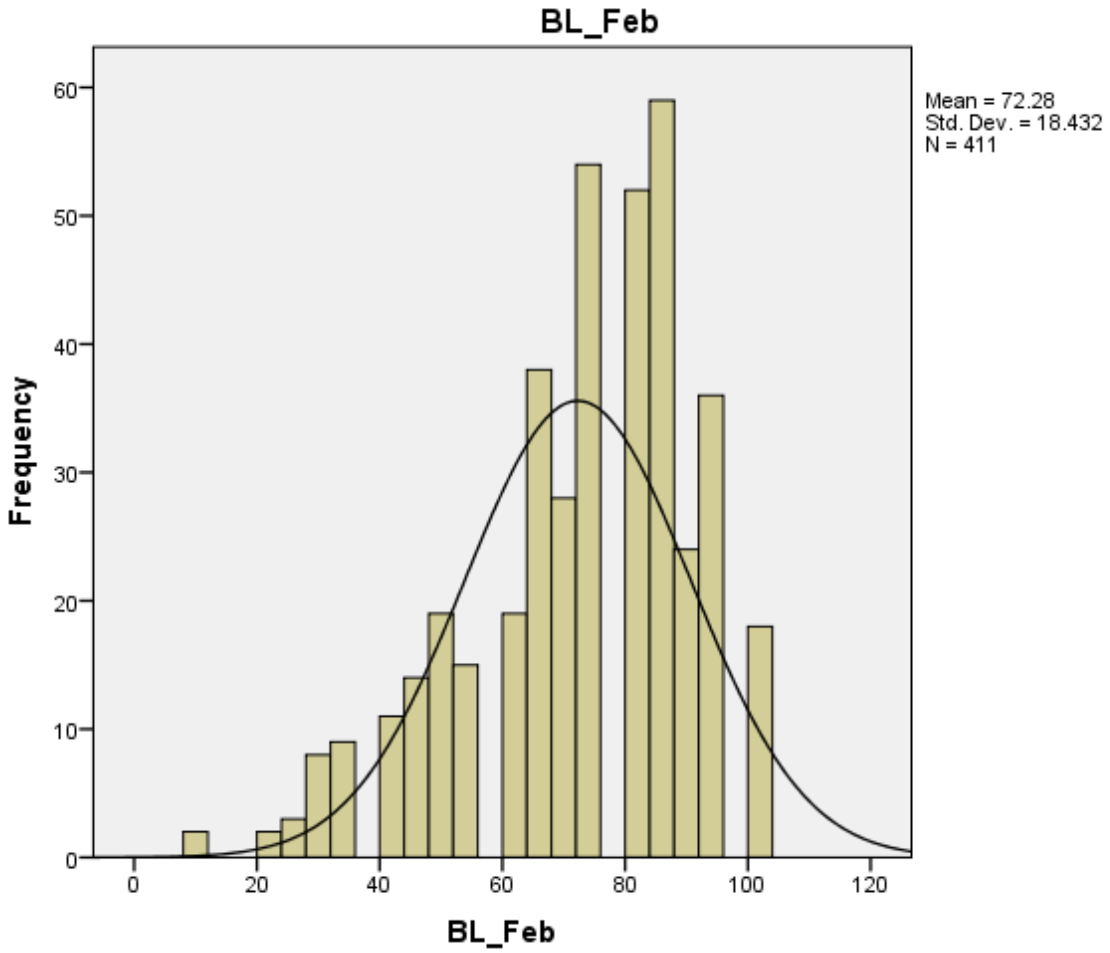


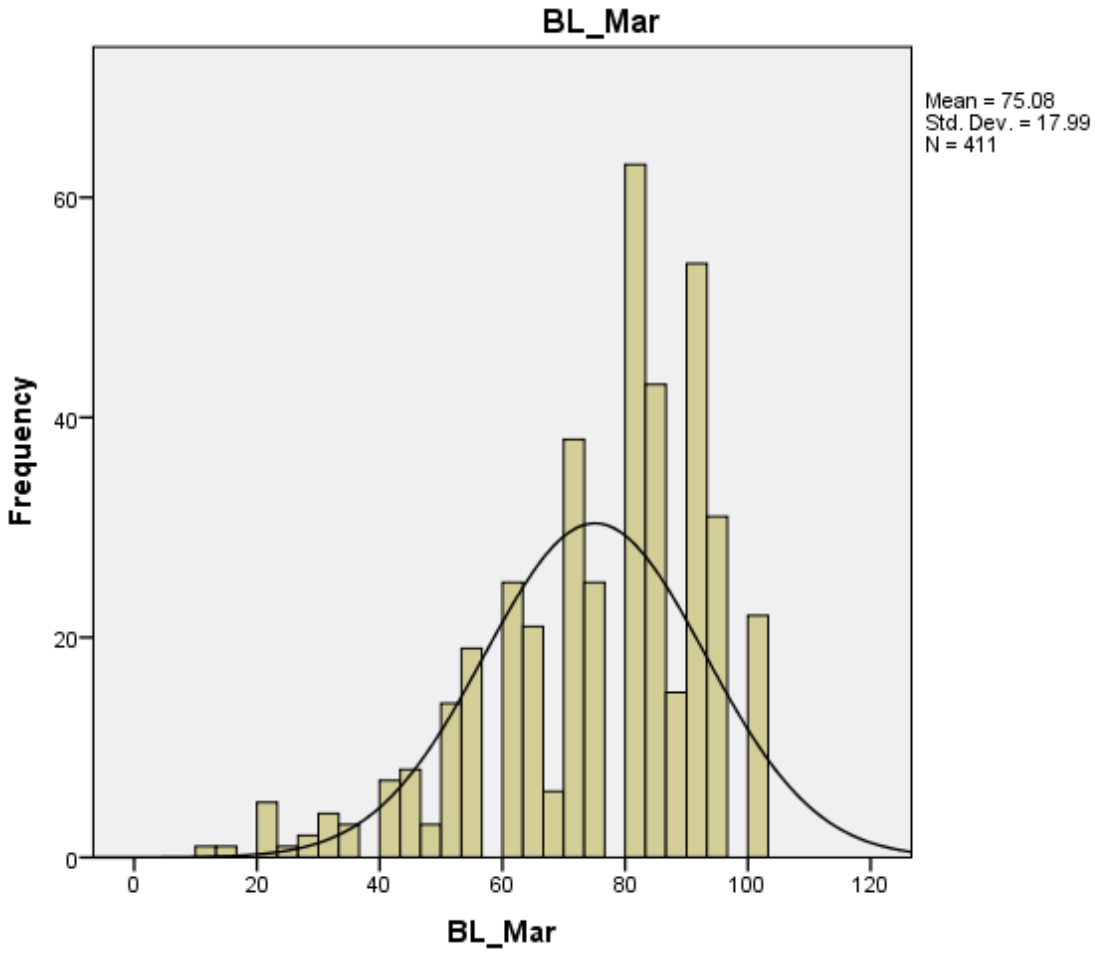

BL_Oct
Mean = 66.16
Std. Dev. = 18.767
N = 411
Frequency
0
10
20
30
40
50
60
0
20
40
60
80
100
120
BL_Oct

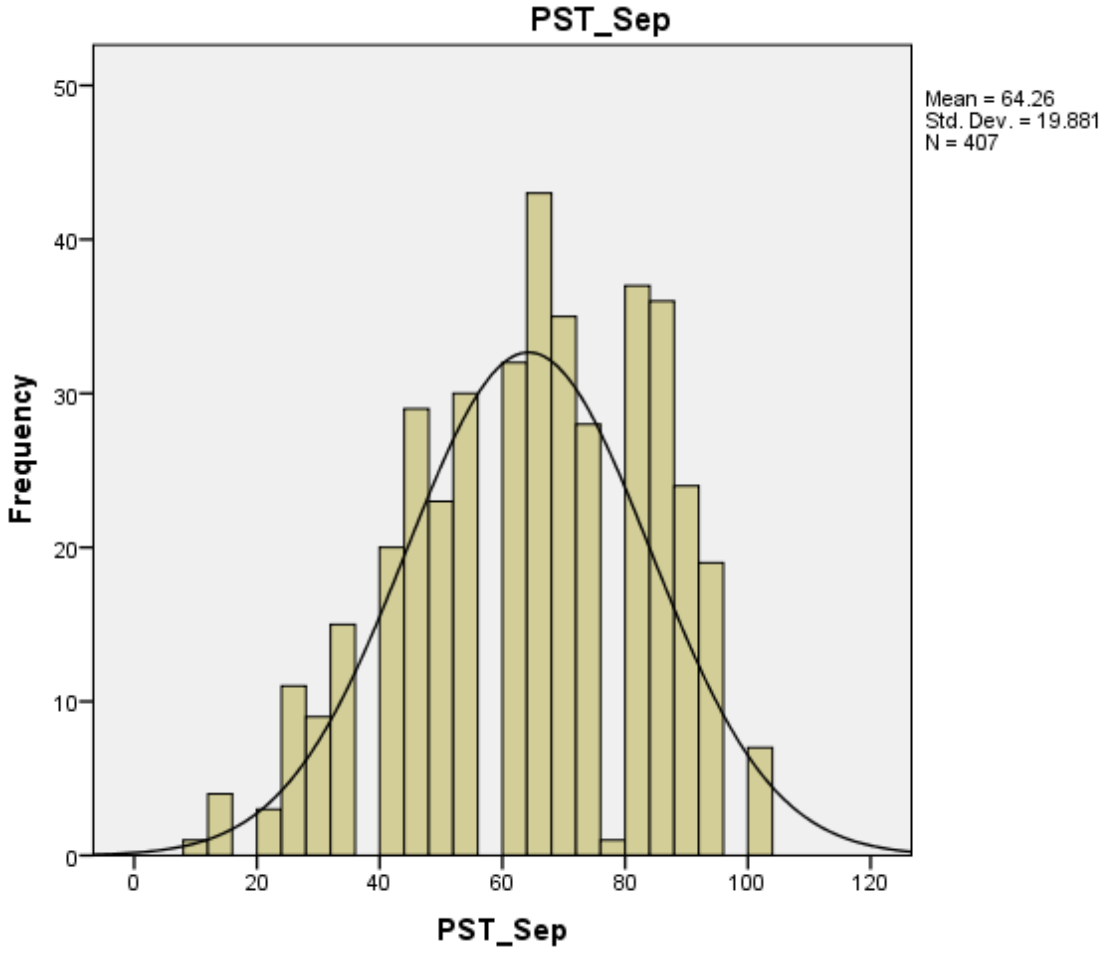

BL_Nov
Mean = 71.35
Std. Dev. = 20.289
N = 411
Frequency
0
10
20
30
40
50
0
20
40
60
80
100
120
BL_Nov

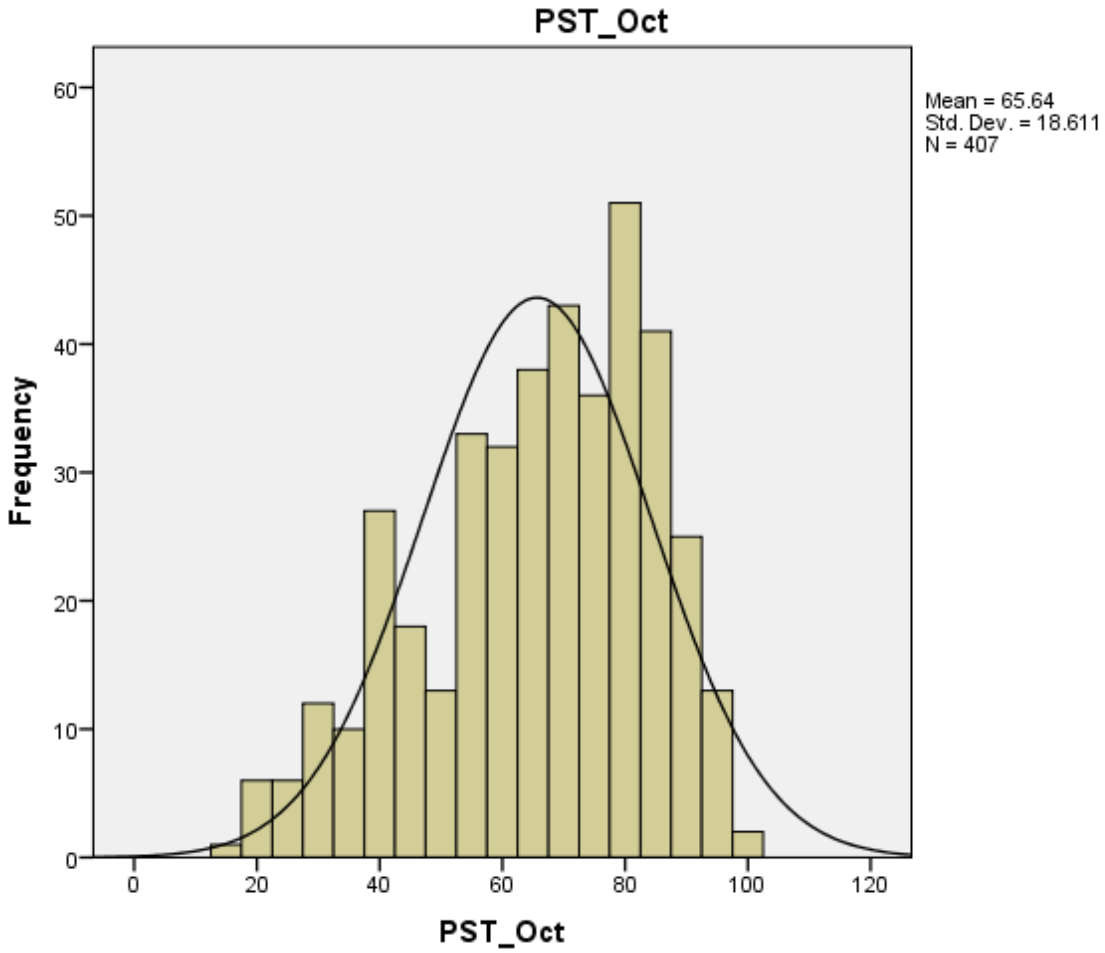

BL_Dec
Mean = 75.7
Std. Dev. = 18.499
N = 411
Frequency
0
10
20
30
40
50
60
0
20
40
60
80
100
120
BL_Dec

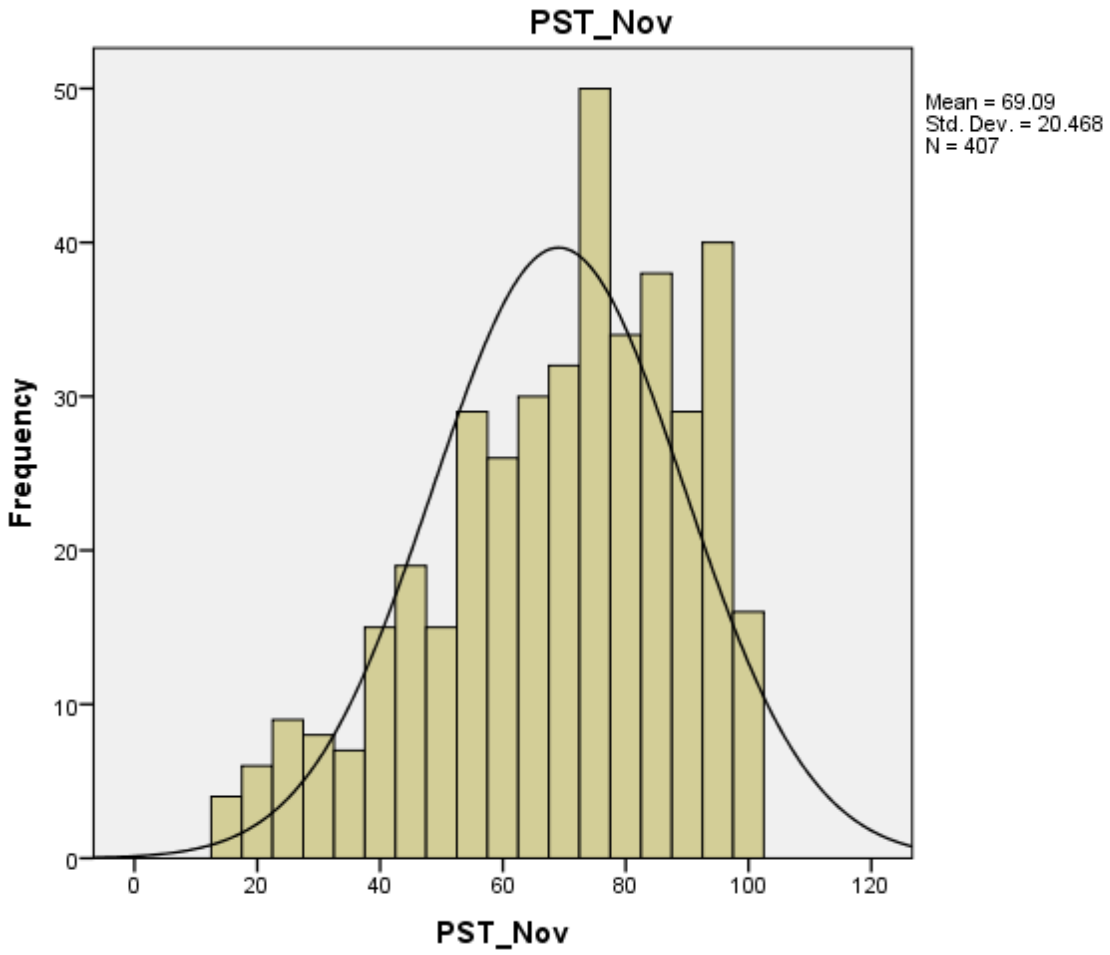

BL_Jan
Mean = 73.44
Std. Dev. = 19.202
N = 411
Frequency
0
10
20
30
40
50
0
20
40
60
80
100
120
BL_Jan

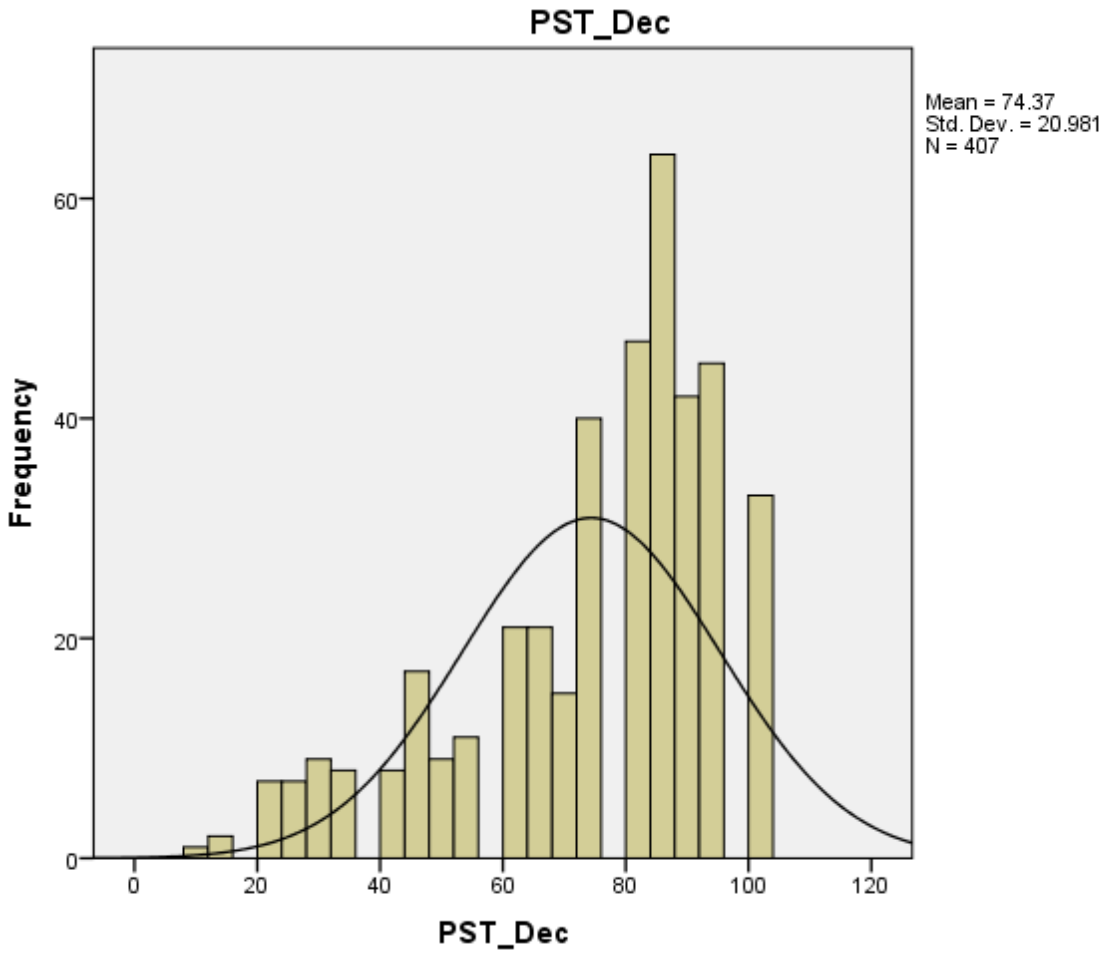

BL_Feb
Mean = 72.28
Std. Dev. = 18.432
N = 411
Frequency
0
10
20
30
40
50
60
0
20
40
60
80
100
120
BL_Feb

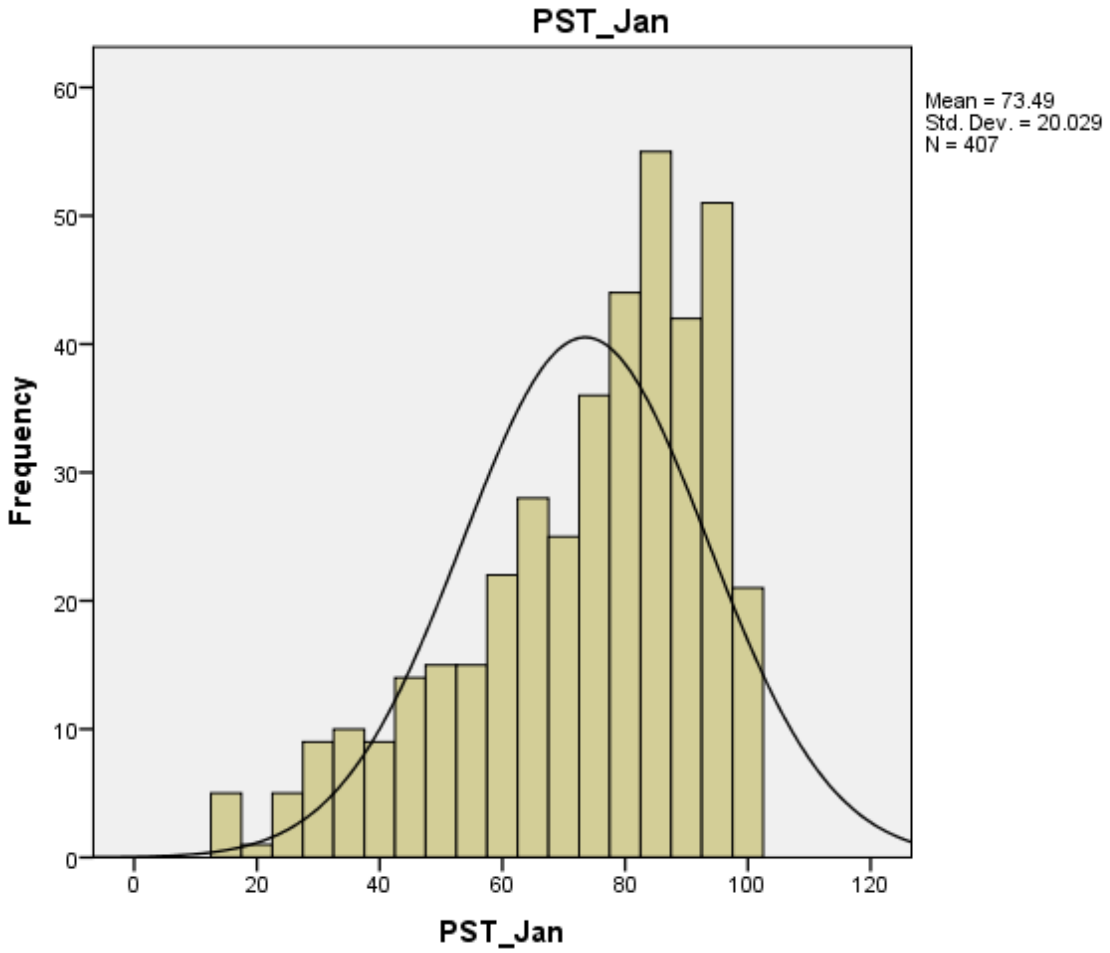

BL_Mar
Mean = 75.08
Std. Dev. = 17.99
N = 411
Frequency
0
20
40
60
0
20
40
60
80
100
120
BL_Mar

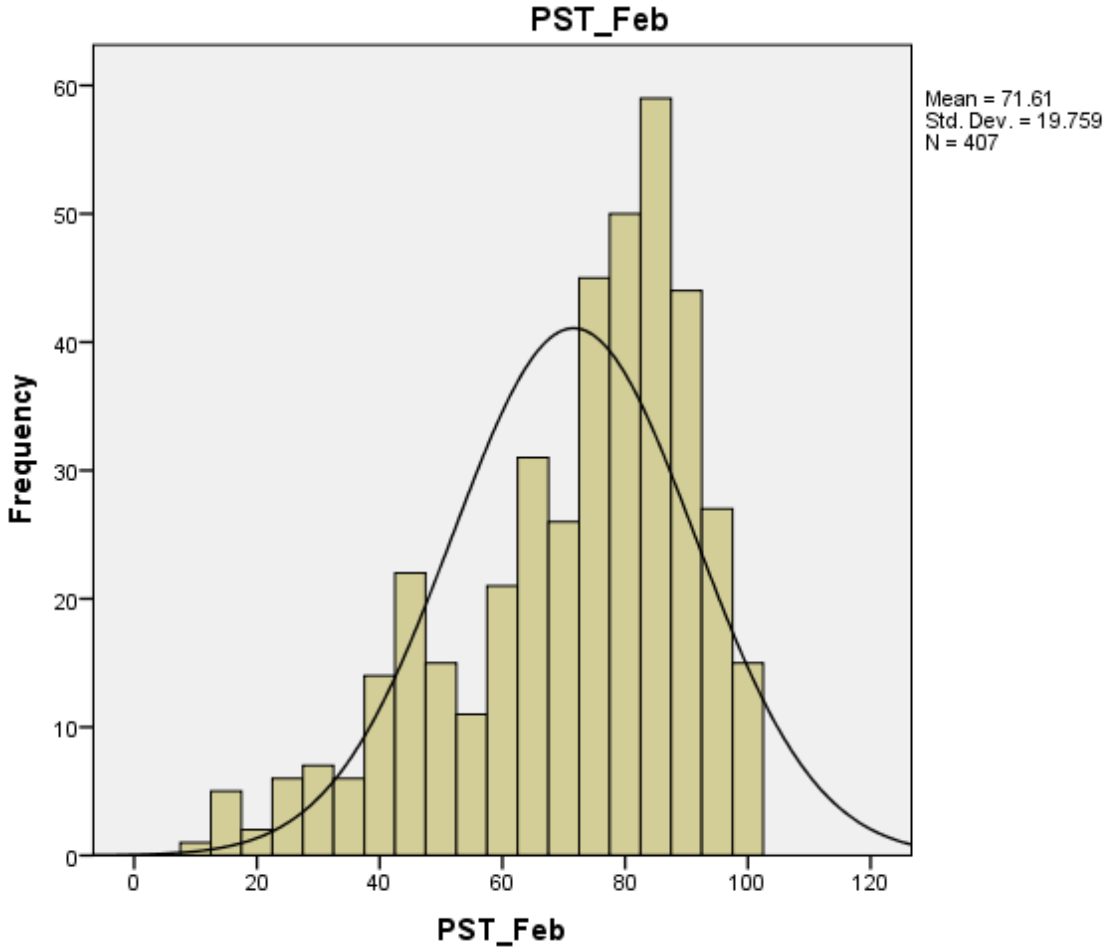

PST_Sep
Mean = 64.26
Std. Dev. = 19.881
N = 407
Frequency
50
40
30
20
10
0
0
20
40
60
80
100
120
PST_Sep


PST_Oct
Mean = 65.64
Std. Dev. = 18.611
N = 407
Frequency
0
10
20
30
40
50
60
0
20
40
60
80
100
120
PST_Oct


PST_Nov
Mean = 69.09
Std. Dev. = 20.468
N = 407
Frequency
0
10
20
30
40
50
0
20
40
60
80
100
120
PST_Nov


PST_Dec
Mean = 74.37
Std. Dev. = 20.981
N = 407
Frequency
0
20
40
60
0
20
40
60
80
100
120
PST_Dec


PST_Jan
Mean = 73.49
Std. Dev. = 20.029
N = 407
Frequency
0
10
20
30
40
50
60
0
20
40
60
80
100
120
PST_Jan


PST_Feb
Mean = 71.61
Std. Dev. = 19.759
N = 407
Frequency
0
10
20
30
40
50
60
0
20
40
60
80
100
120
PST_Feb

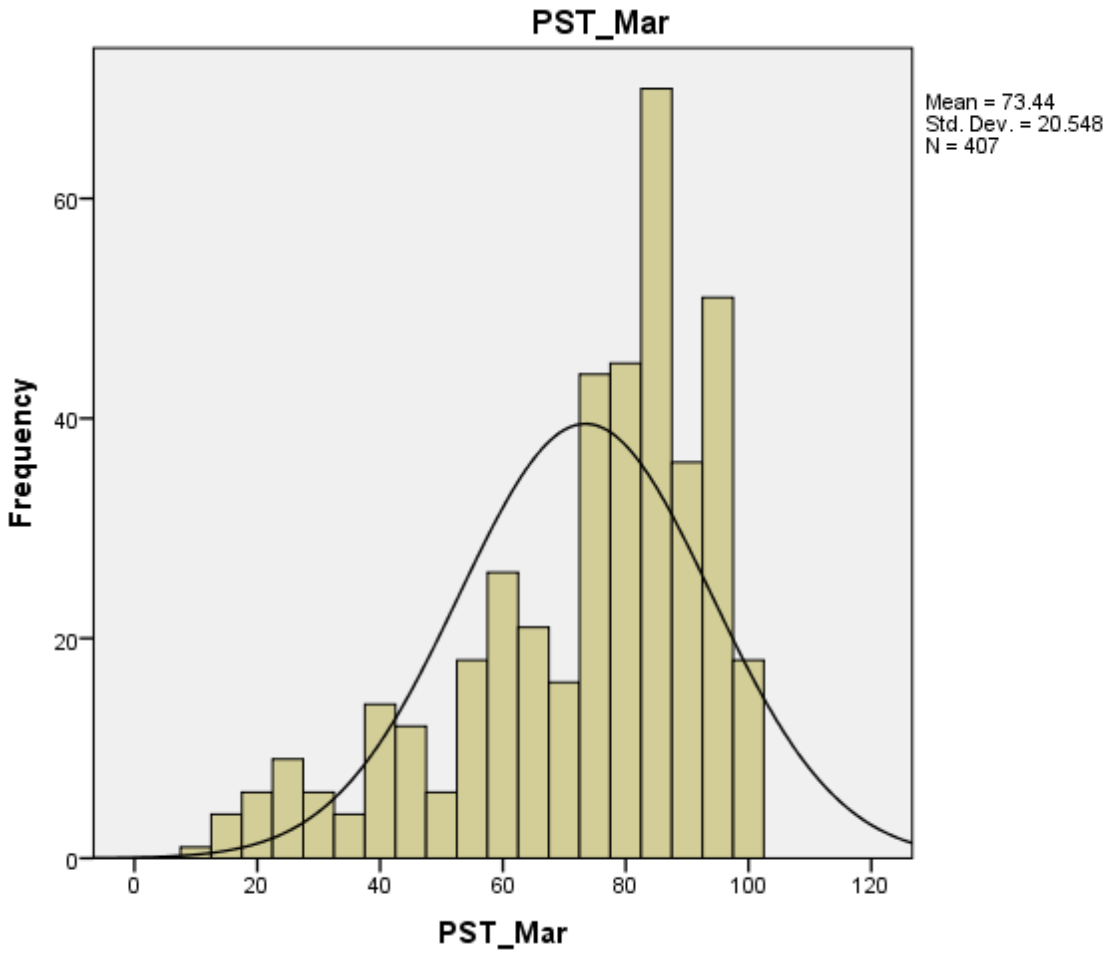

PST_Mar
Mean = 73.44
Std. Dev. = 20.548
N = 407
Frequency
0
20
40
60
0
20
40
60
80
100
120
PST_Mar

Table L1

Demographic Descriptors of Teacher Participants

Teacher descriptors	Control Group	Experimental Group
Average years of teaching experience	13.3	12.9
Percentage of teachers with a masters degree	74.4	78.4
Number of male teachers	1.0	0.0
Number of female teachers	9.0	13.0
Teachers with prior professional Development in essential learning goals	10.0	0.0
Teachers with prior experience in the development of essential learning goals	0.0	0.0

Table L2

Student Demographic Information of Students in Control and Experimental Groups

Student Demographic Indicators	Experimental group	Control group
Total school enrollment	318.0	464.0
Percent asian	3.5	2.3
Percent black	7.3	5.5
Percent hispanic	1.2	1.5
Percent indian	0.0	0.0
Percent white	84.6	87.5
Percent free and reduced lunch	20.2	14.3

Table L3

Mean Scores for Baseline and Post for A_t and B_{nt}

	Group	Mean	Std. Deviation	N
BL_Sep	At	68.80	19.113	85
	Bnt	59.97	18.051	140
	Total	63.31	18.910	225
BL_Oct	At	71.81	16.003	85
	Bnt	64.38	19.294	140
	Total	67.19	18.441	225
BL_Nov	At	80.61	16.398	85
	Bnt	70.24	21.704	140
	Total	74.16	20.458	225
BL_Dec	At	83.22	16.218	85
	Bnt	72.93	19.311	140
	Total	76.82	18.844	225
BL_Jan	At	79.44	15.942	85
	Bnt	71.21	19.017	140
	Total	74.32	18.322	225
BL_Feb	At	76.28	16.875	85
	Bnt	68.60	17.303	140
	Total	71.50	17.507	225
BL_Mar	At	82.62	14.333	85
	Bnt	75.06	18.282	140
	Total	77.92	17.262	225
PST_Sep	At	69.62	18.778	85
	Bnt	61.71	19.953	140
	Total	64.70	19.851	225
PST_Oct	At	69.69	18.993	85
	Bnt	64.04	19.379	140
	Total	66.17	19.387	225
PST_Nov	At	72.73	18.416	85
	Bnt	66.46	20.419	140
	Total	68.83	19.879	225
PST_Dec	At	78.18	18.042	85
	Bnt	69.57	23.534	140
	Total	72.82	21.982	225

	At	80.86	16.094	85
PST_Jan	Bnt	66.89	22.395	140
	Total	72.17	21.316	225
	At	77.49	16.243	85
PST_Feb	Bnt	64.69	21.933	140
	Total	69.52	20.885	225
	At	82.19	14.438	85
PST_Mar	Bnt	63.99	23.245	140
	Total	70.87	22.173	225

Table L4

Adjusted Means and Standard Error for Teacher Experience

Group	Time	Covariate Adjusted Mean	Covariate Adjusted Standard Error
At	BL-Sep	67.566	2.068
	BL-Oct	71.833	2.052
	BL-Nov	80.759	2.249
	BL-Dec	83.16	2.061
	BL-Jan	80.031	2.023
	BL-Feb	75.36	1.928
	BL-Mar	81.608	1.898
	PST-Sep	69.007	2.204
	PST-Oct	68.932	2.169
	PST-Nov	71.921	2.220
	PST-Dec	78.525	2.447
	PST-Jan	79.84	2.279
	PST-Feb	76.926	2.258
	PST-Mar	80.546	2.273
Bnt	BL-Sep	60.721	1.587
	BL-Oct	64.366	1.574
	BL-Nov	70.154	1.726
	BL-Dec	72.967	1.581
	BL-Jan	70.846	1.552
	BL-Feb	69.16	1.479
	BL-Mar	75.674	1.456
	PST-Sep	62.089	1.691
	PST-Oct	64.499	1.664
	PST-Nov	66.955	1.703
	PST-Dec	69.36	1.877
	PST-Jan	67.511	1.749
	PST-Feb	65.03	1.732
	PST-Mar	64.99	1.744

Table L5

Adjusted Means and Standard Error for Education in Years

Group	Time	Covariate Adjusted Mean	Covariate Adjusted Standard Error
At	BL-Sep	69.561	2.136
	BL-Oct	72.313	2.100
	BL-Nov	80.201	2.304
	BL-Dec	81.826	2.095
	BL-Jan	78.092	2.062
	BL-Feb	76.701	1.987
	BL-Mar	83.427	1.954
	PST-Sep	69.783	2.264
	PST-Oct	69.795	2.231
	PST-Nov	73.566	2.278
	PST-Dec	78.973	2.504
	PST-Jan	82.232	2.334
	PST-Feb	76.967	2.315
	PST-Mar	82.865	2.360
Bnt	BL-Sep	59.509	1.624
	BL-Oct	64.074	1.597
	BL-Nov	70.492	1.752
	BL-Dec	73.777	1.593
	BL-Jan	72.022	1.568
	BL-Feb	68.346	1.511
	BL-Mar	74.569	1.486
	PST-Sep	61.617	1.721
	PST-Oct	63.975	1.696
	PST-Nov	65.957	1.732
	PST-Dec	69.088	1.904
	PST-Jan	66.059	1.775
	PST-Feb	65.006	1.760
	PST-Mar	63.582	1.795

Table 6

Descriptive Statistics of Baseline Mean Scores for A_t and B_{nt} Combines School

Assessment Month	n	Min	Max	Mean	Standard Deviation	Skewness		Kurtosis	
	Stat	Stat	Stat	Stat	Stat	Stat	Std Error	Stat	Std Error
September	411	10	100	63.47	18.538	-.453	.120	-.395	.240
October	411	15	100	71.35	18.767	-.454	.120	-.493	.240
November	411	10	100	66.16	20.289	-.665	.120	-.221	.240
December	411	15	100	75.70	18.499	-.939	.120	.343	.240
January	411	15	100	73.44	19.202	-.775	.120	-.111	.240
February	411	10	100	72.28	18.432	-.810	.120	.226	.240
March	411	13	100	75.05	17.990	-.969	.120	.768	.240

Table L7

Baseline Mean Scores for Groups At and Bnt

Descriptive Statistics

	Group	Mean	Std. Deviation	N
BL_Sep	At	67.84	17.690	179
	Bnt	60.10	18.510	232
	Total	63.47	18.538	411
BL_Oct	At	70.32	17.088	179
	Bnt	62.94	19.397	232
	Total	66.16	18.767	411
BL_Nov	At	76.36	17.813	179
	Bnt	67.47	21.250	232
	Total	71.35	20.289	411
BL_Dec	At	79.92	16.497	179
	Bnt	72.44	19.319	232
	Total	75.70	18.499	411
BL_Jan	At	78.42	16.908	179
	Bnt	69.59	20.002	232
	Total	73.44	19.202	411
BL_Feb	At	76.65	17.374	179
	Bnt	68.91	18.552	232
	Total	72.28	18.432	411
BL_Mar	At	79.79	14.622	179
	Bnt	71.44	19.473	232
	Total	75.08	17.990	411

Table L8

Teaching Experience in Years for Group A_t and B_{nt}

PST_Exper

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.0	40	6.7	9.8	9.8
	4.0	30	5.1	7.4	17.2
	5.0	22	3.7	5.4	22.6
	7.0	75	12.6	18.4	41.0
	8.0	28	4.7	6.9	47.9
	9.0	77	13.0	18.9	66.8
	10.0	22	3.7	5.4	72.2
	12.0	23	3.9	5.7	77.9
	16.0	17	2.9	4.2	82.1
	17.0	10	1.7	2.5	84.5
	23.0	24	4.0	5.9	90.4
	24.0	20	3.4	4.9	95.3
	27.0	19	3.2	4.7	100.0
	Total	407	68.6	100.0	
Missing	999.0	186	31.4		
Total		593	100.0		

Table L9

Education in Years Beyond Bachelors

PST_EdYears

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.0	23	3.9	5.7	5.7
	1.0	17	2.9	4.2	9.8
	2.0	176	29.7	43.2	53.1
	4.0	120	20.2	29.5	82.6
	4.5	19	3.2	4.7	87.2
	6.0	28	4.7	6.9	94.1
	8.0	24	4.0	5.9	100.0
	Total	407	68.6	100.0	
Missing	999.0	186	31.4		
Total		593	100.0		

Table L10

Descriptive Statistics of Posttest Scores for the 2011 – 2012 School Year

Assessment Month	n	Min	Max	Mean	Standard Deviation	Skewness		Kurtosis	
	Stat	Stat	Stat	Stat	Stat	Stat	Std Error	Stat	Std Error
September	407	10	100	64.26	18.538	-.453	.120	-.395	.240
October	407	15	100	65.64	18.767	-.454	.120	-.493	.240
November	407	15	100	69.09	20.289	-.665	.120	-.221	.240
December	407	10	100	74.37	18.499	-.939	.120	.343	.240
January	407	15	100	73.49	19.202	-.775	.120	-.111	.240
February	407	10	100	71.61	18.432	-.810	.120	.226	.240
March	407	10	100	73.44	17.990	-.969	.120	.768	.240

Table L11

Scores Over Time in Baseline Year (2010 – 2012) Between Groups

	Group	Mean	Std. Deviation	N
BL_Sep	At	67.84	17.690	179
	Bnt	60.10	18.510	232
	Total	63.47	18.538	411
BL_Oct	At	70.32	17.088	179
	Bnt	62.94	19.397	232
	Total	66.16	18.767	411
BL_Nov	At	76.36	17.813	179
	Bnt	67.47	21.250	232
	Total	71.35	20.289	411
BL_Dec	At	79.92	16.497	179
	Bnt	72.44	19.319	232
	Total	75.70	18.499	411
BL_Jan	At	78.42	16.908	179
	Bnt	69.59	20.002	232
	Total	73.44	19.202	411
BL_Feb	At	76.65	17.374	179
	Bnt	68.91	18.552	232
	Total	72.28	18.432	411
BL_Mar	At	79.79	14.622	179
	Bnt	71.44	19.473	232
	Total	75.08	17.990	411