Defining teacher quality: An examination of the relationship between measures of teachers' instructional behaviors and measures of their students' academic progress

Pamala J. Carter

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Education.

Chapel Hill
2008

Approved by:
George Noblit
William Sanders
Jim Trier
Thomas James
Lloyd Davis
Dick Corbett

ABSTRACT

Pamala J Carter
Defining teacher quality: An examination of the relationship between measures of teachers' instructional behaviors and measures of their students' academic progress
(under the direction of George W. Noblit)

This study will investigate the varying behaviors associated with efforts by teachers to plan, prepare and implement instructional practices for their students. Efforts will be made to distinguish differences in these behaviors in an effort to identify those behaviors that are associated with teachers that have produced high value-added gains scores among the students they teach.

This study will explore possible measures that might begin to provide such data for improving teacher effectiveness. Findings from this study provide useful information in exploring the complexities involved in observing effective teaching that promotes student learning. There are factors that, in this study that are proven predictors of teacher effect scores. There are other factors, of this study that have proven to make a difference when experience has been considered. There are, no doubt, other factors that are yet to be explored that may contribute as much or more to the equation of "best" instructional behaviors proven to increase student progress.

It is the belief of the researcher that a perfect observational instrument has not been found. Much more work is necessary before we put children in jeopardy of being recipients of poorly constructed instructional strategies that are not grounded in a series

of sound and rigorous scientific inquiries. Without a connection between teaching and learning, the act of teaching becomes negligible. We must continue to ask ourselves if the observational instruments we use explain practices found to promote student learning. It is imperative that educational researchers continue to search and question tools being used to determine instructional practices that are most effective at promoting student learning.

To my mom, Lucille Carter and brother, David L. Carter

ACKNOWLEDGEMENTS

This dissertation is not an end, but a beginning. I am indebted to many for making this journey possible. First, I must thank the teachers of Hamilton County. Not only did they take the time to complete a long survey on their preparation and instructional practices, they allowed me to come into their classrooms and observe their instruction. Without them, this study on effective teaching methods would not be possible.

The members of my committee (George Noblit, Jim Trier, Dick Corbett, Lloyd Davis and William [Bill] Sanders) are due many thanks for the endless drafts and edits they endured. George is the best advisor anyone could ask for. He patiently guided me through the mundane adjustments of the process, but more importantly stretched me intellectually to the extremes that only he can get away with. Conversations with Jim Trier always stretched my thinking and brought up ideas and issues in a different light than one would normally pursue. His guidance in helping me explore the history of teacher effectiveness made the literature review rich and interesting. Dick Corbett kept me true to the classroom teacher and the students. He always helped me keep on track with the recipients of the study. From the beginning of this journey, Dr. Davis took me under his wing and shared his research practices and knowledge with me. Throughout my master's and doctorate programs these skills and many life skills were further enhanced by his patience, wisdom and nurturing spirit. Bill's strong belief in the use of accurate

measurement to evaluate student academic progress transcends the conventions of the average research. His questions did not only push me academically and intellectually. I'll always remember his advice to: (a) test your theory and (b) be sure you are accurate and (c) stick to the story despite its popularity. He constantly reminded me that future generations are dependent upon our research and this matter should not be taken lightly.

Several other individuals are due thanks for providing their support. I also cannot forget the patience and attentiveness of Paul Wright in addressing my often redundant statistical questions. Thanks for taking the time to lead me through the complexities of factor analysis. Dr. June Rivers kept me going during the tough times, providing emotional support, professional counsel and comic relief. Thank you to my friends, Leslie Graitcer, Sunhwa Parks, Sharyn Rosenberg, and Deb Vaughan for lending your ears as I worked through the many edits and word changes.

Important in any project is the visionary and that person for me is Jack Murrah. His vision for this work started me on this journey of examining teacher effectiveness several years ago. He trusted in my intuitive nature and allowed me to use those skills in my first studies of teachers in Hamilton County. He continued to nurture those skills and pushed me to think deeper and delve further into the questions that continually surfaced as we discussed ways to ensure all students were getting the best education possible.

Ruth Holmberg, thank you is not sufficient. I am extremely grateful to you for believing in me and trusting that there was more out there for me, even when I was not sure myself.

The journey would not have been possible without my family. The journey would have been less enjoyable with my partner, Marcy Thayer. Thank you for supporting me as this journey continues.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The importance of the role that teachers play in student achievement and the issue of teacher effectiveness has been a topic of interest to researchers and the educational community for some time. Although stated in various ways, the premise of the message is the same: When it comes to student achievement…teachers matter (Darling-Hammond, 2000; Haycock, 1998). A few large-scale studies have found the teacher to be a major factor in student learning (Jordan, Mendro & Weerasinghe, 1997; Sanders & Rivers, 1996; Wright, Horn & Sanders, 1997). Research studies on this issue are numerous but, unfortunately, replications of such studies, especially on a larger scale, are scarce. Various indicators are used to address the issue. Many different instruments are used to collect the data. The metrics applied to those indicators further complicate matters.

Of critical concern is the use of so many different definitions in determining what it means to be an effective teacher. The search for criteria that will determine characteristics of an effective teacher continues to perplex the educational research community. Although the field of study has dramatically improved, there is work to be done. This work is essential in providing evidence useful in the training and promoting of teachers that provide the most impact on our students' academic achievement.

In this opening chapter, background information, useful in supporting the premise of this study, will be introduced. The purpose in pursuing such an investigation, along with suggestions for methods to employ, is discussed. Value-added models will be

presented as a means for investigating the issue. Promising evidence from previous value-added studies has identified teachers that produce the most student learning, defined by academic achievement gains. A brief argument supporting the study's intended value-added model, Tennessee Value Added Assessment System (TVAAS), will follow. The significance of these findings will, in turn, provide more information that can be applied by the educational community to facilitate improved instruction. This information, including professional development training and preservice programs, will more adequately equip teachers with scientifically-proven measures shown to promote student academic achievement gains.

This study will investigate the varying behaviors associated with efforts by teachers to plan, prepare and implement instructional practices for their students. Through surveys and observation, the researcher will gather data to support the various activities occurring within a classroom setting and the preparation that has taken place (via teacher self-report), prior to that lesson, to determine behaviors associated with the most academic gains with students. A search for a "method of appraisal" as suggested by Gage (1972) that would objectively and accurately measure teacher effectiveness assessed through student learning is a focus of this study. Efforts will be made to distinguish differences in these behaviors in an effort to identify those behaviors that are associated with teachers that have produced high value-added gains scores among the students they teach.

Background

Literature on teaching from the late 1800s suggests the need for examination of teacher effectiveness. In *Theory and Practice of Teaching*, Page (1885) writes:

> During the earlier and the greater part of the historic period, all scholars were necessarily teachers, and it was an easy step to the inference that all who were learned could teach. …as the number of schools greatly increased, it was observed that some scholars had high teaching power, while others had little or none of this gift. As this difference could not be attributed to differences in scholarship, nor wholly to differences in natural ability, it was ascribed to high and low degrees of skill, and so the question of method was called into prominence. This step necessarily led to comparisons of methods and finally to a search for some criterion by which they could be tested (p. 19).

Research on teacher effectiveness, however, has its beginnings in the 1920s with much of the earlier works being framed around the administrator perspective (Dunkin & Biddle, 1974; Gage, 1965). The 1930s and 1940s gave rise to presage-process-product studies where various characteristics of teachers were examined for their relationship with teacher effectiveness on student learning (Campbell, Kyriakides, Muijs, & Robinson, 2003). These studies considered the formative experiences, called presage variables, of the teacher that were present prior to entering the classroom. Variables that represented actions that occurred in the classroom, i.e. activities of the teacher and the students, were called process variables. The product variables were the outcomes that most often related to the progress determined as a result of the process implemented (Dunkin & Biddle, 1974). Experimental studies were popular during the 1940s through the 1960s, including investigations of different teaching styles such as formal and informal and progressive and traditional (Mitzel, 1979).

From the 1960s forward, much of the research attention on teacher effectiveness has concentrated on teacher knowledge and beliefs and their relationships to student progress (Campbell, et al., 2003). Medley and Mitzel (1963) purport that an important factor in determining student learning is the teacher. They further state that factors

attributing to this effectiveness, once they have been identified, can be taught in teacher education programs to enhance the practice of new teachers. A more concrete and measurable definition of effectiveness will enable a program of research and analysis that will lead to a better understanding of what it means to be an effective teacher as well as a better understanding of how to teach those that want to become effective.

Given the urgency for teacher accountability under the current federal administration, the time to address the issue is now. The federal government's intent with its accountability system (within the No Child Left Behind act) is to identify problem areas – areas where all children are not making progress – and provide corrective action for improvement. With this federal call for action, more attention has moved to the investigation of methods and skills employed by teachers who are getting better results. Although multifaceted, these methods and skills of effective teachers are not unattainable. Using the knowledge of the previous decades' research on teacher effectiveness, researchers continue to investigate aspects of the effective teacher that promote student learning. With today's computational and statistical capabilities, not only can these studies serve to answer the call from the federal government for accountability purposes, they can serve the much larger purpose of informing future educators of more effective ways to promote student academic progress.

Purpose of the study

The premise of this study assumes that teaching's purpose is to produce student learning, defined by academic achievement. A scientific approach linking teacher behaviors to student achievement will provide evidence of teaching behaviors that

promote learning, i.e., increased student academic achievement. To date, although there is a growing consensus in the role the teacher plays in student learning, there is no agreement on a single definition of teacher effectiveness. Research has progressed to its current use of student standardized test score gains, relative to a particular teacher's class, to determine teacher effectiveness (Sanders & Horn, 1994; Thum, 2003; Webster & Mendro, 1997). As we seek to determine what makes a teacher effective, now and into the future, I see a value-added dimension of examining student achievement included in the equation.

The purpose of this exploratory research is to investigate effectiveness specifically among math, reading and language arts teachers in grades 3-8, defined by their (a) planning and preparation and (b) instructional practices. Through structured observation and survey assessments, the researcher will examine defined aspects of the teacher to collect data on behaviors associated with each dimension. The Tennessee Value-Added Assessment System (TVAAS) model's teacher effect is the dependent variable defining effectiveness. The researcher will analyze the above data, seeking practices that discriminate effective from ineffective teachers.

Common in current literature is the notion that the single greatest predictor of student learning is the effectiveness of the teacher (Darling-Hammond, 1991, 2000; Ferguson, 1991; Haycock, 1998; Sanders & Horn, 1998; Webster & Mendro, 1997). Proponents of this mindset have suggested the use of a value-added component in evaluation systems for educational assessment (Webster & Mendro, 1997). The use of longitudinal data in determining educational progress is imperative for the future academic growth of our children.

Value-added models are numerous and range from very simplistic calculations of gains scores to very computationally intensive mixed-models. The models are built to chart *value* added within a specific time period, to examine how much improvement is made toward a proficient goal and/or to use as an evaluation tool for reviewing effectiveness of programs and personnel (Doran & Izumi, 2004). The term value-added can be found in 1950s literature in the tax structure of Western European nations and is prominent internationally in agricultural communities (Cowan, 2003; Poulson, 1965). In communicating value-added in these communities, one would use phrases such as "goods, value and wealth in political economy" (Thum, 2003, p. 157). The modern terminology's context can be better described as "relative progress," coined by Goldstein and Spiegelhalter (Thum, 2003). Several defining terms have been used in an attempt to create a politically correct and acceptable definition.

In education, the term value-added can be redefined to include the use of longitudinal test data to measure adjusted comparisons of student data and its changes over time (Doran & Izumi, 2004). Students enter the classroom at different academic achievement levels. Value-added is able to utilize the level of achievement upon entry and compare that level to attainment at various stages over time. These adjusted comparisons provide representation of how much *value* has been added to a student's learning. That is, instead of measuring student achievement by examining student scores at a single point in time, researchers are able to look at individual scores as they relate to that individual's previous scores (Sanders & Horn, 1994). "A value-added model seeks to measure teachers' influence on student learning and achievement gains by comparing students' yearly performance while also using additional descriptive or confounding

variables to increase the explanatory power of the model" (Germuth, 2003, p. 14). Information provided by the value-added model is useful in evaluating teacher performance.

Utilizing value-added models, researchers have begun to discriminate the patterns of teacher effectiveness and have found that some teachers are consistently better at producing higher gains than others (McCaffrey, Lockwood, Koretz & Hamilton, 2003; Thum, 1997; Sanders & Horn, 1998; Webster & Mendro, 1997). These patterns have raised questions about the possibility of being able to improve student learning by studying specific cases. If there are differences among teachers that produce higher gains than their counterparts who do not and these differences can be proven, documented and more importantly, duplicated, the impact on student learning would be substantial.

In several states, through complex statistical value-added models, we are now able to adequately determine the amount of individual student learning that can be attributed to a teacher in a given year (McCaffrey, et al., 2003; Thum, 1997; Sanders & Horn, 1998; Webster & Mendro, 1997). The statistical technique employed through value-added methods is capable of isolating and examining the effects on student achievement from year to year. This technique allows one to connect teachers with the achievement of their students more directly. It can provide meaningful information regarding how teachers impact a variety of students, relative to their prior performance, and may be used for continuous improvement.

These value-added models provide an opportunity for researchers to reexamine classrooms for behaviors and instructional practices that are linked with teachers found most effective. With this knowledge, we can examine teachers that are more effective at

promoting student learning, determine what they are doing, and report that information for the professional learning/growth of others that might not be as effective. To examine the usefulness of value-added in determining a more accurate picture of accountability, the most popular model to date, the Tennessee Value-Added Assessment System, will be discussed (Thum, 2003).

The TVAAS model, created by Dr. William Sanders, has the benefit of over twenty years of longitudinal data to support its methodology and provides the much-needed empirically-based dependent variable, the teacher effect. The model has removed student background variables arguing that in using the student as his own control or blocking factor, a better estimate of teacher effects on student achievement is possible. Creators of the model suggest that by including background variables, areas that should be addressed will be masked (Sanders & Horn, 1994). Prior learning of each individual is measured and growth is determined from this previous learning, not on an average of the learning of other students that might have similar economic and ethnic backgrounds (Sanders & Horn, 1994).

Developed in the 1980s, TVAAS is the most widely known and used model of value-added in education. Sanders, Saxton & Horn (1997) write that it provides a "best linear unbiased estimate when examining the influence of student achievement gains attributable to teachers, schools and school districts" (Millman, 1997, p. 139). Millman (1997) states, "The mixed-model approach is far more complex and less well known [than is regression]. It is familiar to many statisticians studying plant and animal breeding and other agricultural applications, but it is little-known to the human behavior

statisticians" (p. 167). Tennessee and, more recently, Ohio have adopted this accountability system at a state level (Ceperly & Reel, 1997; SAS Institute, 2008).

Successfully following students as individuals requires a tremendous effort on the part of schools, districts and states. A student must be given a unique identifier and that identifier must follow him throughout his schooling. With this unique identifier, each student can be followed even if she changes schools or districts within the state. This unique identifier allows researchers to collect the needed longitudinal data to fit into the value-added model to determine the amount of gain from year to year of any given student. The Tennessee Value-Added Assessment System has this capacity (Jordan, Mendro & Weersinghe, 1997). The model requires three years of continuous data. The model is viewed as robust, fair, reliable and valid as a statistical value-added method especially when using longitudinal data and the complexities attached to such data (Stronge and Tucker, 2000).

Given the current situation of national educational research, teacher accountability has become more important than ever. The now-mandated standardized test given annually to all students is an attempt by the federal government to ensure an equitable academic learning environment for all children. States are being held responsible for educating their children. The current research on the teacher as the single most important determinant of student achievement has pushed the accountability of this learning ultimately to the students' teachers. The urgency of moving student learning forward has pushed the need to find practices that are effective at helping teachers accomplish this task.

Significance of the study

This study can have an impact on the educational community immediately and in the future**.** Areas that could benefit include teacher preparation academies, state, district and local professional development programs and public policy issues of teacher performance, quality and distribution. The ultimate goal would be to identify teacher practices that are instrumental in positively affecting student academic achievement.

This study will benefit both teachers and the educational community. It will contribute to what we know about raising student achievement and how best to accomplish it. The ability to delineate effective from ineffective practices will inform teacher preparation programs. Effective practices that make a difference in student academic achievement can be brought to the teacher education colleges and become part of the coursework required for future teachers.

With proven effective practices, not only will teacher preparation programs benefit, but, also, school districts can capitalize on the findings. Professional development programs can be designed to improve teaching. Incorporating effective practices into professional development programs will strengthen the current teaching workforce, thus improving student academic growth.

Findings from this study can address educational policy. Given the charge of meeting the goals of No Child Left Behind, states are continually searching for ways to improve the quality of instruction, thus improving student achievement in their districts. Providing districts/states with findings that are shown empirically to positively relate specific practices to improved student achievement could address policy areas such as teacher pay for performance, redistribution of the teaching workforce and objective

teacher evaluation. Several districts and states have already adopted a pay for performance policy. These states are providing incentives to teachers that "perform" above what is expected. The ability to provide practices that are positively linked with this higher performance will present policymakers with evidence on which to base merit pay increases.

The ability to strengthen planning and preparation practices using the results from this study also is possible. If specific planning and preparation practices are found to be more effective at raising student academic achievement, teachers benefit from this information and can immediately take this information back to the classroom.

Instructional practices found positively related to increased academic gains can benefit teachers both in the preservice programs and the classroom. Effective practices can be observed, studied and modeled by teachers found to be effective at moving students forward. The findings from this study and evidence of successful instructional practices will be presented as areas to consider for future preservice programs and district professional development opportunities.

Decisions are continually made in an effort to provide more successful educational systems for students. Better decisions can be made given a more reliable measure of determining practices that are most effective in raising student academic achievement. This study will explore possible measures that might begin to provide such data for improving teacher effectiveness.

Research question

This investigation will assess the (a) planning and preparation and (b) instructional practices of math, language arts and/or reading teachers for grades three through eight. This study will seek to respond to any distinguishable differences found among these teachers. The study will utilize the TVAAS model to categorize teachers with varying levels of effectiveness. The research question: What are the differences in (a) planning and preparation and (b) the implementation of instructional practices of teachers at varying levels of effectiveness? This question will be addressed using math, reading and/or language arts classes in grades three through eight in one school district.

CHAPTER II

LITERATURE REVIEW

The educational research community has not been idle in its attempts to respond to the need for effective teachers across the century. In this chapter, a review of the literature will present the reader with information on teacher effectiveness research including (a) a synopsis of previous research beginning in the early 1900s through the 1960s; (b) some of the seminal research from the 1960s through the present; (c) models of teacher effectiveness; (d) current research on teacher effectiveness; (e) measuring teacher effectiveness through student achievement; and (e) the methodological challenges confronting the uses of the current complex methods.

Equally important to the review of teacher effectiveness research studies is the methodology used in these studies. An overview of observational systems used within educational research will be discussed. As well as reviewing the various instruments used by observers, surveys designed to capture the teacher perspective of behaviors within the classroom also will be discussed. These sections of the literature review will culminate with the framing of the current study.

Introduction

There is debate over what effective teaching should look like and what characteristics an effective teacher should possess. The conversation grows more

complex as the multitude of perspectives joins the dialogue. In the 1998 summer issue of *Thinking K-16,* Kati Haycock of the Education Trust reminds us that good teaching matters and it matters a lot. Most would not disagree with that statement. What *is* effective teaching? How do we define the characteristics? Many such questions are raised as researchers examine this issue.

Much of the research on teacher effectiveness over the course of the last century is consistent from decade to decade. A primary assumption of these efforts is that the quality of instruction does differ among teachers and that researchers can identify what accounts for these differences (Medley, 1987). Cruickshank (1990) noted two distinct eras when reviewing the history of the research on teaching, pre- and post-1960s. Most research students of the pre-1960s era focused on identifying the characteristics or traits that a teacher possesses after having been identified as a good teacher by his administrator or supervisor. During the 1960s, a different approach began to emerge as researchers examined a teacher's ability to promote student learning. Teachers were seen as an important influence on student achievement (Rosenshine, 1970). Conversations during this time addressed the notion of effectiveness not only as a research question but as a means to inform and improve teacher education programs. Since the 1960s, much of the research on teaching examines teacher behaviors as they relate to student academic success. A closer examination of some of the most influential studies of these two periods informs the reader of the complex issues that surround and confound the educational research community.

Pre-1960

Studies and research on teaching during the early 1900s examined the teacher, with little attention given to the act of teaching or the learning interactions that might have occurred in the classroom. Much of the research focused on the relationship of the personal characteristics and qualities of the teacher with his perceived ability to teach. Some of these characteristics and qualities include enthusiasm, compassion for children and knowledge of the content. The assumption was that models of teaching behavior would provide information to the educational community without having to literally investigate the learning process. These results, however, had little impact on teaching (Bellon, Bellon, & Blank, 1992).

These early studies were not exempt from the complexities involved in investigating characteristics of effective teachers. Beecher (1949) provides a thorough review of teacher research beginning with the teacher evaluation work of J.L Merriam in 1905 through the Wisconsin studies in the mid-1940s (Barr, Worchester, Abell, Beecher, Jensen, Peronto, Ringness, & Schmid, 1961). Merriam's work was credited with "taking the problem of teaching efficiency from the field of opinion…and placing it in the field of research and objective measurement" (Beecher, 1949, p. 5). A.S. Barr (1929) and his investigation of good and poor teachers and Bryan's (1937) rating scale of secondary school teachers represent some of the various types of studies coming out of this period.

Merriam (1905) examined 1185 graduates of normal school training programs and tried to show the relationship between their schooling and teaching ability. Although he had no accurate measure for teaching ability or a way to measure environmental factors, the study is credited as a "well conducted pioneering experiment" (Beecher, 1949, p. 5).

He found very low correlations between the “scholarship” of their schooling and their teaching practice. Some of the correlations between the normal school achievement and teaching ability following graduation found most interesting by Merriam include “practice teaching (0.39), psychology (0.37), history and principles of education (0.28), methods courses (0.29) and academic courses (0.22)” (Beecher, 1949, p. 6).

A. S. Barr (1929) compiled a list of teacher evaluations in order of frequency as given by 106 school superintendents. He gathered names of good and poor teachers and checked their ratings with state inspectors. Those with the highest and lowest ratings participated in the study. He divided the 47 teachers in grades seven through twelve in the study into two groups according to their teaching ability, superior and less than average ability. Examiners studied good teaching to discover common behaviors. Some qualities studied included teacher actions, vocabulary used, comments to pupils, materials used and the organization of the materials and lesson. Teaching also was examined to determine commonalities among the teachers as well as differences among the poor teaching behaviors observed. Through observation, Barr determined that the differences on the ratings among the highest and lowest ranking teachers were not significant. He also found that the data gathering procedures, the measures being used and that the terminology used among supervisors was highly subjective and unreliable. Qualitative and quantitative analysis concluded that factors should be reviewed for their appropriateness to the context when considering them as factors related to teacher effectiveness (Barr, 1929).

Bryan’s (1937) Pupil Rating of Secondary School Teachers rating scale was an 11-item questionnaire on pupil opinion of the teacher. Teacher ratings were collected

from 1500 students and their perspective teacher ratings were collected from administrators. The purpose of the study was to "determine the reliability of pupil ratings of teachers and the degree of agreement between pupils' and administrators' ratings of teachers and also to determine what items in the rating instrument have most weight in determining general teaching ability" (Beecher, 1949, p. 17). Significant findings from Bryan's study reported that five items in these ratings were found significant in determining teaching ability. They were "(a) the amount pupils are learning; (b) the amount of work the teacher does; (c) sympathy; (d) the ability to explain clearly; and (e) knowledge of subject" (Beecher, 1949, p. 19).

The reliance on checklists, rating systems and teacher records of the 1930s and 1940s began to decrease and the need for more accurate measures to determine characteristics of master teachers emerged. It was during the mid-1940s that researchers began to investigate the relationship of pupil growth with teacher effectiveness.

In 1953, a Committee on the Criteria of Teacher Effectiveness presented the grim description of the current research of teacher effectiveness as not responsive to nor answering the needs of the modern American education system (Gage, 1972). Five years later, Brim's (1958) review of the literature showed no consistent relations between teacher characteristics and the effectiveness of teaching, measured by the academic success of that teachers' students. Wallen and Travers (1963) echoed the sentiment reporting that various teaching methods and how they related to teacher effectiveness made little difference and one method could not be favored over another. Even with the reported inadequacies in some of the research that was being conducted, the need for

more information on the effective teacher kept researchers moving forward in their search.

The 1960s saw a shift from the personal characteristics of the teacher to a focus on teacher-student interactions. Personal characteristics previously found were often viewed as innate and thus hard to alter. The focus on effective teacher behaviors suggested that, if identified, these behaviors could be taught to prospective teachers (Bellon, Bellon & Blank, 1992).

The study of teacher effectiveness since the 1960s

Most studies of the 1960s continued to support the idea that nothing within a teacher's instruction seemed to make a difference in the level of effectiveness, measured in terms of student academic achievement. A study completed by Barr and his associates (1961) from the University of Wisconsin, in collaboration with the Wisconsin State Department of Public Instruction, sought to determine which selected measures brought about specific changes in students. In the study:

> …attempts [were made] to determine the relationship between selected teacher measures and changes produced in pupils by these teachers….At the same time, the experiment seeks to find what combination of measures gives the highest correlation with teaching ability as measured by the criterion of pupil change (Beecher, 1949, p. 25).

Data collected included test measures from both pupil and teacher. The study was extensive. Students were given pre- and post-tests at the beginning and the end of the school year, along with various mental and reading tests. Teachers were subjected to a battery of twelve tests including the American Council Psychological Examination, Yeager's Scale for Measuring Attitude toward Teachers and Teaching Profession,

Bernreuter's Personality Inventory and Toregerson's Test of Teaching Problems and three rating scales measuring various traits such as enthusiasm and resourcefulness. A key assumption of the study was "measurable differences in pupil learning could be accepted as criteria for teaching ability" (Beecher, 1949, p. 25). General conclusions emerging from this multifaceted study reported that: "(a) intelligence seems to be the highest conditioning factor in teacher success of those checked; (b) social attitudes are important; (c) attitude toward teachers and teaching is important; (d) knowledge of subject and ability to diagnose pupil maladjustments are important; (e) correlations between supervisor ratings and criteria of the study, namely pupil change, were insignificant; and (f) personality, as defined, is insignificant" (Beecher, 1949, p. 27).

Gage (1965), recognizing the inadequacies in the research on teacher effectiveness suggested a continued search for an answer in that the "need for knowledge in this area is pressing" (Gage in Flanders & Simon, 1969, p. 1424). He concluded that from previous research five characteristics "seem" to be components found in effective teachers. These components include the (a) warmth, (b) cognitive organization, (c) orderliness, (d) indirectness, and (e) problem-solving ability of the teacher (Gage, 1965).

The research of Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld & York (1966) reported that the social and family background contributions, along with the attitudes that students brought to the classroom accounted for most of the variance in student achievement, not the teacher or the instruction. Stephens (1967) concurred as he referenced a host of studies that investigated various educational variables and found that none had a significant and consistent impact on teacher effectiveness.

Gagne (1970) built on the theory of Stephens and suggested that perhaps variables could be more accurately measured with a different method. He suggested that the variables should be considered as process variables and thus be viewed as "human actions which transform the raw materials of input into opportunities for learning" (Gagne, p. 170). These teacher activities would be examined instead of the previous characteristics such as education, experience and verbal ability. The research community began to review the studies that had been conducted on these process variables with a different lens. Contrary to Coleman and his colleagues (1966) and Stephens (1967), maybe there was something in a teacher's instruction that made a difference in their effectiveness.

The research methodology of the 1970s began to address this issue. By using the individual teacher as the unit of analyses instead of groups of teachers or the school, researchers now were able to differentiate behaviors among teachers. Using student achievement as the criterion for success, these methodologies were able to differentiate between more and less effective teachers. Using results that delineated the more effective behaviors, teachers could be given specific guidance on areas for improving their instruction.

Soar (1971) collected observational data on seventy classrooms. This study revealed that "there seemed to be a trend that abstract growth was related to teaching that was less controlled and less focused but had some structure, while skill growth was related to more focus and structure, with concrete growth positively related to still more highly focused teacher behaviors and negatively related to extreme pupil freedom" (Soar, 1971, p. 1). These findings also were supported by Brophy & Evertson (1976). Brophy &

Evertson (1976) studied 28 teachers over the course of two years. During the first year, teachers were observed about 10 hours and about 30 hours during the second year. Through the extensive and comprehensive research design, they collected much data on the teachers. Both low and high inference measures were collected, combining the questionnaires and interviews for a total of 580 presage variables to be used in the analysis. Rosenshine and Furst (1973) found the use of a wide variety of questions by the teacher to be a crucial factor in their research from the 1960s and the early 1970s.

In a 1972 publication, Gage used the Teacher Characteristic Schedule developed by Ryans (1960). His study involved 6000 teachers in 1700 schools within 450 school systems. The study "dealt with relationships between estimates of teacher-behavior patterns observed in the classroom, inventory of estimated teacher characteristics, background and environmental variables, and observed pupil behaviors" (Biddle & Elena, 1964, p. 67). The instrument had fairly consistent positive correlations with observer ratings of elementary teachers and student scores on achievement tests. Variables excluding pupil behavior were measured by self-reporting instruments; the pupil behavior by direct observation. "His design is classical in the sense that teacher 'characteristics' are abstracted from the classroom context. His work is unique in that he has established relationships between the characteristics described and both formative and outcome variables" (Biddle & Ellena, 1964, p. 32).

Brophy and Evertson (1976) conducted the Texas Teacher Effectiveness Project. They began their study with the intent to measure "anything" they thought might correlate with student gains. After reviewing the literature on teacher effectiveness measures, they built "a single integrated and complex system that would take into

account as many variables as possible, rather than [to use] a number of different systems that overlapped considerably and that would have reduced the data available on each variable" (Brophy & Everston, 1976, p. 174). They observed nineteen second and third grade teachers for two consecutive years and found the stability of classroom behaviors across the two years varied greatly. Efforts to explain this variance included factors such as the reliability of the instruments used to collect the data, the frequency of the behaviors observed and the context within which the behavior was observed (Brophy & Evertson, 1976).

In 1979, Good and Grouws conducted an experiment on a mathematics teaching program, the Missouri Mathematics Effectiveness Study. Of the 40 teachers in the study, one group was trained using the effective direct instruction behaviors and the other group was not given the training and taught as they had before. Their observational measures revealed that the students of teachers trained to implement the program outperformed students from the control group. They found a significant difference in the progress of the students taught by the teachers of the experimental group. They concluded that more effective teachers (a) taught the class basically as a whole; (b) presented information more clearly; (c) were task-oriented; (d) created relaxed learning environments; (e) had higher achievement expectations; and (f) had fewer discipline problems (Good & Grouws, 1979). This study was supported by a similar study in the subject of English (Fitzpatrick, 1982). Fitzpatrick (1982) found that behaviors could be taught to teachers and that those exhibiting these behaviors were found to be more able to engage the students in their classroom (Muijs & Reynolds, 1999).

The work of the 1960s and 1970s served as evidence to support teacher-centered instruction that dominated the 1980s (Darling-Hammond & Wise, 1985; Wilson & Corbett, 1990). The shift of the research began to move from the former context of teaching as presented in the studies of the prior two decades to the importance of the learning that was or was not occurring in the classroom. The patterns of behaviors that had been found among the teachers were being studied within the context of the lesson. This research provided the educational community with a deeper understanding of the pedagogical knowledge which, in turn, strengthened the understanding of principles and theories that explained teacher effects on student learning (Porter & Brophy,1988).

Good, Brophy and Biddle reported in 1975 that 10,000 studies had been conducted on the topic of teacher effectiveness. Rosenshine (1973) concurred and remarked to the small number of correlational and experimental studies conducted that tested the relationship between classroom events and student outcomes. Those with the most consistent results reported clarity; variability; enthusiasm; task-oriented behavior; criticism; teacher indirectness; student opportunity to learn criterion material; and the use of structuring comments and multiple levels of questions as being positively related to student outcomes (Rosenshine, 1973). These results were similar to findings from the studies of the 1960s and early 1970s (Biddle & Ellena, 1967; Brophy & Evertson, 1976; Gage, 1965; Soar, 1971).

The imperative of increasing student academic standards came out of the Elementary and Secondary Education Act of the mid-1960s and was in full swing in the 1980s. Accountability for the monies that had been funneled to the state came into question and a call went out for more testing of both students and teachers. Inevitably,

teachers became a central component in the accountability movement. In 1983 with the report, *A Nation at Risk*, there was a push for a return to the "basics" of education to provide the essentials that many felt were lacking with the current educational system. The need to determine the most effective methods of teaching was necessary to move the nation forward. To find the most effective methods, researchers turned to schools to determine what characteristics were present in the most effective teachers within these schools.

Good (1983) reported that no single instructional strategy had been found to be unvaryingly successful as a means of teacher effectiveness. He suggested that teachers that skillfully used a range of approaches were most often more successful. Vogt (1984) went a step further and proposed that effective teaching required attention to be given to the varying abilities of the different students. He suggested a ranking system of gauging the teacher as exceeding, meeting, needing improvement and unsatisfactory. These rankings gauged the teacher's ability to incorporate the instructional objectives and to assess the learning modes of different students (Vogt, 1984).

In the mid-1980s, several researchers reviewed studies that examined relationships between various school factors and student achievement. Hanusek (1986) reported that most of the empirical studies of this time had limited sets of tools and most analysis was based on single regression analysis to estimate the relationship between these factors and student achievement. He further concluded that such analyses resulted in skewed and biased evaluations. Raudenbush and Bryk (1986) concurred with Hanusek on the need for multilevel instead of single regression analyses. They reported that the majority of studies of educational effects actually "conceal more than they reveal"

(Raudenbush & Bryk, 1986, p. 1). The issue of measuring the relationship between student achievement and teacher effectiveness required a more complex statistical analysis than had ever been used in educational research (Raudenbush & Bryk, 1986).

Instruments of the 1990s included those that recorded various behavior patterns of student-teacher interaction. Studies during this time examined, coded and tallied behavioral patterns observed during classroom instruction. Behaviors and interactions included lesson clarity, flexibility of instruction, pacing of the lesson and the levels of conversations occurring in the classroom (Berliner & Biddle, 1995; Brown & Wagg, 1993; Chuska, 1995; Cruickshank & Metcalf, 1994). Researchers during this time were busy working to create more effective models for capturing these behavioral patterns and instructional practices.

Findings from these studies gave us the many characteristics we still study as best practices such as: problem-solving ability of the teacher (Curwin & Mendler, 1997 & 1999; Slavin, 1997); classroom climate (Bamburg, 1994; Brophy, 1992; Cotton, 1997; Creemers & Reezigt, 1997; Fraser & Walberg, 1991; Freiberg, Stein & Huang, 1995; Good & Brophy, 1997; Muijs & Reynolds, 1999; Wang, Haertel & Walberg, 1993); having clarity (Borich, 1996; Brophy, 1992; Cantrell, 1998/1999; Good & Brophy, 1997; Muijs & Reynolds, 1999; Taylor, Pearson, Clark & Walpole, 1999; Teddlie & Stringfield, 1993); variability/flexibility (Evertson, Emmer, Sanford, Clements & Worsham, 1997); enthusiasm (Cabello & Terrell, 1994; Good & Brophy, 1997; Newby, 1991); the use of students' prior knowledge (Adams, 1990; Ball, 1997; Bransford, Brown & Cocking, 1999; Hunter, 1994; Lambert & McCombs, 1998; Putnam & Borko, 1997); the use of structuring comments and multiple levels of questions (Chuska, 1995; Gall &

Gall, 1990; Goodlad, Soder & Sirodsnik, 1990; Moore, 1995; Rosenshine & Meister, 1992; Wilen, 1991); high achievement expectations (Good & Brophy, 1997; Jones, 1990; Weinstein & McKown, 1998); evaluation and feedback (Borich, 1996; Brualdi, 1998); and classroom management skills (Burke, 1997; Cotton, 1996; Evertson, Emmer, Sanford, Clements & Worsham, 1997; Jones, 1990; Muijs & Reynolds, 1999). The research community has progressed in its empirically-based classroom studies that link teacher practice to student achievement. Still, they continue to struggle to define teacher effectiveness, how best to identify the "best" practices and how best to measure qualities of the "effective" teacher. There remains little consensus on these issues.

Current research on teacher effectiveness

As researchers continue to examine the relationship between student achievement and teacher practice, many have declared teacher effectiveness to be the single largest determinant of student achievement (Darling-Hammond, 2000; Goldhaber & Brewer, 2001; Hanusek, Kain, & Rivkin, 1998; Mosteller, Light & Sachs, 1996; Sanders & Rivers, 1996; Schmidt, McKnight, & Raizen, 1997; Wenglinsky, 2000; Wright, Horn & Sanders, 1997). Mosteller, Light and Sachs (1996) are known for their examination of the effects of class size from Tennessee's Project Student-Teacher Achievement Ratio (STAR) study. Concluding that teachers could increase student achievement given smaller class size, this study was seminal in President Clinton's decision to push for legislation to support smaller class sizes.

Sanders & Rivers (1996) reported on the cumulative effects of the sequence of teachers a student might have. In analyzing data from a cohort of 2nd through 5th grade

students from the Tennessee state database, they calculated teacher effect scores and reviewed the distribution of teachers within the various quintiles. Individual student records were then matched to the corresponding teachers. Results concluded that students, beginning at the same level of academic achievement, tend to respond similarly to teacher effectiveness levels (Sanders & Rivers, 1996).

Schmidt, McKnight and Raizens (1997) Third International Mathematics and Science Study (TIMSS) provided evidence of differences in student achievement among various countries investigated and explained these differences with the process-product effects of the teacher. Wright, Horn and Sanders (1997) considered the size of teacher effects on student achievement when examining academic growth. They considered not only the teacher effects, but other influences such as intraclass heterogeneity, the academic level of the student and the relationship of class size to academic growth. The results of this study revealed that "teacher effects are dominant factors affecting student academic gain and that the classroom context variables of heterogeneity among students and class sizes have relatively little influence on academic gain" (Wright, Horn & Sanders, 1997, p. 57).

In 2000, Darling-Hammond reiterated and reported the importance of the teacher and how their effectiveness outweighs class heterogeneity and class size. Hanusek, Kain, and Rivkin (1998) report teacher quality as an important determinant of school quality. They suggest that the effects on achievement differences found in schools can be attributed to the variation in teacher quality. They report that the variations in teacher quality account for at least 7.5% of the total variation in student achievement. "Even if teachers were randomly distributed among schools (highly unlikely) and all of the

between school variation in achievement were to result from other school inputs (it is even more unlikely that students are randomly distributed among schools), differences in teacher quality would swamp all other school inputs" (Hanusek, Kain & Rivkin, 1998, pp. 30-31).

Wenglinksy (2000) built on the work of Wright, Horn and Sanders (1997) using NAEP data to identify practices that improved the student outcomes from an 8th grade science report. He reported that teacher input, professional development and classroom practices were all influential in producing high student achievement with classroom practices, especially with those using higher-order thinking strategies being the most significant. Also utilizing NAEP data and the 1993-94 Schools and Staffing Survey Data, Darling-Hammond (2000) researched teacher effectiveness from a different angle. She found that states that invested heavily in improvements to the quality of its teachers and in student accountability showed the highest gains on the NAEP assessments and focused her results on the importance of teacher licensure programs. A third NAEP review of 1988 data by Goldhaber and Brewer (2001) utilized mathematics data to examine the effects of various covariates on student learning and found that teachers with mathematics training outperformed those that lacked such training (Rice, 2003).

These studies that support the claim of the importance of the teacher in determining the academic success of the student capture the multifaceted aspects of undertaking such a complex topic. Methodological challenges exist around every corner. The need for further explanation of the appropriate class size, the true variation accounted for by the teacher, heterogeneity among the students, and other issues continue to push researchers to seek out ways to improve current and future research on teacher

effectiveness. Given the urgency of the topic and the imperative need of a meaningful method, it is appropriate to discuss issues involved in measuring teacher effectiveness through student achievement.

Models of teacher effectiveness

Numerous models of teaching exist, no doubt due to the complexity of the act of teaching itself. These models investigated teacher behaviors, their experiences, the quality of their instruction, properties of the student or some combination of these and other aspects of the teacher and his students, the classroom and teaching itself. (Biddle & Ellena, 1964**,** Carroll, 1963; Collins, 1990; Cruickshank, 1990; Danielson, 1996; Dunkin & Biddle, 1974; McBer, 2000; Shanoski & Hranitz, 1992).

Biddle & Ellena (1964) offered a seven-variable model for the investigation of teacher effectiveness. In this model, (a) formative experiences, (b) teacher properties, (c) teacher behaviors, (d) immediate effects, and (e) long-term consequences serve as main sequences variables; (f) classroom situations and (g) school and community contexts serve as contextual variables (Gage, 1972). Each of the five main sequence variables is postulated to affect the one immediately behind it in sequence. In addition, classroom situations (which are somewhat under teacher control) and school and community contexts (which are not) jointly constrain and interact with the linear effects of the five main sequence variables listed above. …No attempt has been made to state explicit hypotheses for the model" (Biddle & Ellena, 1964, p. 18). The complexities involved in investigating teacher effectiveness are quite apparent in the process described above as well as others that followed.

Carroll (1963) presented a “comprehensive model of factors affecting school learning in terms that provided for (a) quality of instruction; (b) aptitude, expressed as the amount of time required by the learner to attain a specified criterion; (c) perseverance, measured by the amount of time the learner is willing to spend at learning; (d) opportunity to learn, defined as amount of time actually allowed for learning in the particular setting; and (e) ability to comprehend instruction, or perhaps verbal intelligence” (Gage, 1972, p. 95). The statistical methods used to analyze the factors could be used to compare one type of instruction to another (Gage, 1972).

Nearly a decade later, Dunkin and Biddle (1974) suggested that an effective model of teaching would need to concern itself with the properties of both teacher and pupil; characteristics of the classroom; outcomes of education; and the process of teaching itself. In this model they present the behaviors of both the teacher and the student in the classroom as well as observable changes in pupil behavior as a result of either the teacher or the students. The model adequately captures the essence of what is to be observed and presents a concise and often cited diagram designed to orient the reader with areas of concern when addressing a study of teaching. Authors do not delve into analysis procedures however, but spend a large amount of time discussing relevant literature on various research studies that further justify their model (Dunkin & Biddle, 1974). The complexities and sophistication of instruments remained an imperative issue in any model that will undertake the measuring of teacher effectiveness in the classroom.

Two researchers stood out in the early 1990s as important to the literature for their models of teacher effectiveness. First, Cruickshank (1990) designed a “cluster of teacher effectiveness” to include seven traits that were to be accounted for when observing for

effectiveness. The seven traits were (a) teacher character traits; (b) what the teacher knows; (c) what the teacher teaches; (d) what the teacher expects; (e) how the teacher manages; (f) how the teacher reacts to pupils; and (g) how the teacher manages the classroom. Collins (1990) established criteria for an effective teacher while working on his Teacher Assessment Project. Collin's reported an effective teacher as (a) committed to students and learning; (b) knows the subject matter; (c) responsible for managing students; (d) can think systematically about their own practice; and (e) a member of the learning community (Collins, 1990). A couple years after these studies, Shanoski and Hranitz (1992) used Cruickshank's seven clusters to examine effectiveness and supported his findings.

Charlotte Danielson's *A Framework for Teaching* (1996) presented a model of components to be considered in the practice of effective teaching. She presented four major domains of (a) planning and preparation, (b) the classroom environment, (c) instruction, and (d) professional responsibilities. She further explained what each of these domains would look like in an effective classroom. Although presented as a model for evaluation, researchers have begun to use this framework's domains to study teacher effectiveness. Danielson & McGreal (2000) remarks,

> Some educators equate teacher evaluation with classroom observation; others equate it with the forms used. Revising their system of evaluation, then, becomes a matter of changing the forms, or the forms used in an observation. Although evaluation forms are important in defining the structure of an evaluation process and the types of professional conversation surrounding it, forms do not constitute the system. An effective teacher evaluation system is far more complex than the forms and must contain three elements: (a) a coherent definition of the domain of teaching (the 'What' and 'How good is good enough?'); (b) techniques and procedures for assessing all aspects of teaching (the 'How'); and (c) trained evaluators (p. 21).

This statement resonates equally as imperative for use as a research model for teacher effectiveness. The realization that evaluation systems are more complex than the forms used to capture the information should not be taken lightly. The domains of Danielson's framework will be modified and used within the observational system of the proposed study, keeping in mind the complexities described by Danielson in the above quote.

McBer's (2000) model of teacher effectiveness was commissioned by England's Department for Education and Employment. Although only a "short version" of the report has been published, the key findings present the characteristics identified as effective teaching practices by McBer. According to the report, the framework was created from analysis of historical documents of teacher qualifications, skills and characteristics, classroom observations, interviews, questionnaires, focus groups, teacher personal data and school data. In the findings, McBer (2000) reports that there are three main factors influencing student progress that are within the control of a teacher. These factors are (a) teaching skills, (b) professional characteristics and (c) the classroom climate. He suggests that teaching skills and professional characteristics are factors that the teacher brings to the profession while the third factor, classroom climate, is an output measure that deals with the ability of the teacher to understand and work with students in the classroom, providing the motivation that is needed for learning to occur. The study used pre and post tests based on a value-added dimension not described. Although the methodology and analyses are not presented in the report, McBer (2000) purports,

> Using the knowledge and outcomes from our research, we have been able to *model* the impact teachers have on the classroom climate, how that climate affects pupil progress and what aspects of teaching skills and behavioral characteristics had most impact on climate. Our findings suggest that, taken together, teaching skills, professional characteristics

> and classroom climate will predict well over 30% of the variance in pupil progress (pp. 10-11).

Each domain is further described in detail, presenting "micro-behaviors" that are exhibited for each. For example, there are 35 behaviors that are presented as behaviors under the teaching skills domain (McBer, 2000). A modification of the model's framework, using the three domains presented above, will be used in the proposed study.

The models discussed above provide evidence of the complexities involved in measuring teacher effectiveness. The importance of finding a model that adequately addresses these complexities, including the universe of classroom behaviors and learning as expressed through student achievement is imperative to future research. No longer are the rating scales and models that were subject to administrator competence and bias acceptable. The question of what makes a teacher effective has yet to be answered. However, if the motive is the success of the student and that success is to be determined by how well they present their learning outcomes on an annual standardized exam, then a model that considers student-tested learning a function of teacher effectiveness is appropriate.

Measuring teacher effectiveness through student achievement

Much of the current research examining teacher effectiveness has taken a quantitative approach. Hanushek (1986) reviewed the relationship between school factors and student achievement in 147 empirical studies and found that most were based on single equation regression analysis, resulting in biased evaluations and skewed data. He, along with Raudenbush and Bryk (1986) and Stevens, Estrada, and Parks (2000), found

these current accountability models to be inappropriate and labeled them as approaches that produced erroneous results. These findings led to further inquiry into which unit should be analyzed, the student, the classroom or the school. From this inquiry came multilevel, mixed effects or covariance models, also known as hierarchical linear models (HLM), in educational research.

Goldstein (2001), on his belief of the place for HLM in educational research, remarked: “The statistical models now available, together with the powerful and flexible software, enable researchers to explore the inherently complex structure of schooling in a manner that begins to match that complexity” (p. 18). Although adequate methods are still being debated, the HLM models provide better results than evaluation methods that often are subjective and have little, if any connection with student learning (Goldstein, 2001).

Hierarchical linear modeling has its roots in Henderson’s statistical modeling (McLean, Sanders & Stroup, 1991). Raudenbush and Bryk (1986) and McLean, Sanders and Stroup (1991) were among the first to use Henderson’s model in educational research. This approach became a flexible statistical tool that could be used in studying how variations of reform and practice are influential in the educational process. William Sanders was the pioneer in the use of HLM modeling to measure student achievement gains. His model measures student achievement over time on the Tennessee state test and the data obtained is used to measure how effectively a school and/or teacher increases student test scores, i.e. the students’ knowledge (Sanders & Horn, 1994). Sanders coined this process as “value-added” (Sanders & Horn, 1998). This model was adopted by the state of Tennessee and has been used as part of the statewide assessment since the 1980s.

The Dallas Value-Added Accountability System (DVAAS), created by Webster and Mendro (Millman, 1997) employs a slightly different approach using HLM. In a report on DVAAS, Webster, Mendro, Orsak and Weerasinghe (1998) provided a discussion and summary of the statistical approaches used for estimating school and teacher effects from the previous ten years. They suggested that the methodology of choice for producing school estimates is a two-stage, two-level student-school HLM model while that for producing estimates of teacher effect is a two-stage, two-level student-teacher HLM model. From their review, they suggested that most of the methodological issues that have been raised from the literature can now be resolved (Webster, et al., 1997). The model is regarded as commendable, but Thum and Bryk (Millman, 1997) caution about the complexities and inherent problems of the DVAAS and other models that utilize test data to evaluate and measure student outcomes.

Sanders' model (TVAAS), just as the Dallas model (DVAAS), is not without its multitude of critics. Baker and Xu (1995) foresaw the problem of variations of the test scores from year to year along with problems with accurately identifying teacher factors. Sanders answered Baker and Xu stating that his model accounted for their concerns (Millman, 1997). Bock, Wolfe and Fisher (1996) raised concerns over the use of national norms in assessing Tennessee students, along with issues of the model's complexity causing problems with using it as an assessment tool. However, the authors conclude that the central concept of Sanders' model is "the only present, fair, objective and dependable method of evaluating teacher effectiveness based on scores…" (Bock, Wolfe and Fisher (1996, p. 69).

As researchers began using multilevel modeling in examining student gains, several impediments surfaced (Lockwood & McCaffrey, 2007). Not all students took the required standardized tests every year in every subject. By not having complete data on a student, researchers were faced with the issue of what to do with these missing items. It is known that the students that miss the most school are generally the most economically disadvantaged and often minority (Sanders & Horn, 1998). Not including these students in the analysis would produce inaccurate reports on the effectiveness of a teacher, school and/or district.

If a method of imputing data is used, this information will usually come from an average of the group examined. A statistical rule allows the researcher to "impute" or credit a score where no score is actually given. This can be successfully done in some analyses. In the case of analyzing student test data, a student that has no test score will be credited with the average score of her peers. Students are often misrepresented with this method. To get around this problem, many researchers will exclude these students with missing data, opening up the analysis for criticisms on the accuracy of the reporting. TVAAS utilizes all scores available on a child to create as accurate a description of that individual child as possible. Without imputing data, the TVAAS model has found a way to use all relevant data available on a student. This means that all students are included in the analysis, thus a more accurate picture is created (Sanders & Horn, 1998).

An obstacle that continues to perplex researchers is whether to include socioeconomic and ethnic background information in the model. Coleman and his colleagues had reported in 1966 that "pupil achievement could not be significantly elevated until conditions governed by race, class, and income inequality were rearranged

to strengthen the positive role of healthy families" (Fallon, 2004, p. 2). From their findings, Coleman reported that only a small portion of student achievement is a result of school factors, with most of the variance being explained in the family background of the student (Fallon, 2004). Following Coleman's lead, Goldstein (2001) affirmed this finding but also reported that other factors influence student gains, including other teachers (teacher peers), student background and the school setting. He asserted the difficulty of suggesting the "progress of any one pupil in a given subject to the teacher of that subject" (Goldstein, 2001, p. 4). Sanders reports that a better estimate of teacher effects on student achievement is achieved when students serve as their own control, making variables such as socioeconomic status and ethnicity unnecessary. If each child is followed using his own data from previous years, the opportunities for drastic differences in these variables is unlikely. The TVAAS model promotes the omission of these variables (Sanders, Saxton & Horn, in Millman, 1997). Sanders has reported that this model can include socioeconomic variables but to include these would not provide the most accurate finding (Sanders, personal communication, May 12, 2006). The model requires three years of continuous data be used to prepare a strong estimate of gain. The thoroughness and the robustness of this model has proven it to be the most effective model to use in determining effects of district, school and teachers on student achievement (Sanders & Horn, 1998; Stone, 1999).

Many researchers are cautious and skeptical of using student gains to measure the effectiveness of teachers. "Even examples of the best teaching may not provide a theoretical basis for the most effective teaching" (Flanders & Simon, 1969, p. 1424). The concerns from these researchers are not without warrant. Most of the literature on teacher

effectiveness comes from the extensive body of literature on teacher evaluation. Most teacher evaluations are the result of observation (Danielson & McGreal, 2000). This observational information then is crafted into formal rating systems that are used in research studies to determine effectiveness. These rating and ranking systems are met with apprehension when used as evaluation instruments to define effectiveness. As evaluations, they were met with criticisms of subjectivity and bias.

When addressing teacher effectiveness with a value-added measure to define that effectiveness or examining classroom practices for teacher improvement and/or evaluation, the systems used to collect the data need to be carefully scrutinized. A closer examination of the research on educational observational systems as well as a brief review of research on teacher perceptions [via survey response] is warranted prior to the investigation of classroom practices.

Using observations and surveys to measure teacher effectiveness

Observational systems

Boehm and Weinberg (1997) report objective observational techniques as being "central to the scientist's methods of inquiry for generating hypothesis, for building laws of science and for confirming theories" (p. 6). Many of the current observational techniques originated in the 1920s as a result of the creation of the committee of child development by the American National Research Council. The earliest of techniques involved mainly diaries and narrative logs of events. These observational practices of studying the behavior of both men and animals thrived in the 1920s and early 1930s (Hutt & Hutt, 1970). Boehm and Weinberg (1997) further explain that these techniques are not

without their difficulties. Earlier education observational studies spent much energy focusing on the role of schooling and the process as a whole (Foster, Gomm and Hammersley, 1996). Objective observational techniques have become a part of the focus of academic research into the school life and processes of schooling only over the last thirty to forty years (Foster, 1996). Prior to this time, much of the observational study of education was directed to the inputs into and the outputs from schooling but little attention was given to what was going on inside the classroom. The research of the last few decades has seen a radical change and current trends follow a closer examination of happenings in the classroom (Foster, 1996).

The physical nature of the classroom presents complexities that can easily become impediments in observing certain events as they occur. The task of observing and monitoring everything that happens in a classroom is all but impossible (Hook, 1981). "The process of selecting behaviors to be recorded is essentially one of identifying a limited range of behavior relevant to the purpose of the study and of constructing categories or items to be used by the observer" (Medley and Mitzel, 1963). For many studies of educational settings, the current literature on best practices drives the observational plan, particularly if the researcher is looking to identify important gaps in a specific aspect of schooling (Foster, 1996).

Given a focus on what areas of the educational setting are to be examined, the researcher must determine the best mode for collecting the data that has been determined important for the study. For studies with this type of structured process, clear decisions need to be made prior to collecting data to ensure consistency and standardization in the process. Included in these decisions are the option of recording data by frequency and/or

the duration of the behavior being observed (Foster, 1996). Another way of stating this, "The purpose of the observation must guide the selection of the instrument" (Keeves, 1988, p. 473). This instrument would need to be scrutinized to ensure that it is capable of collecting the data that has been deemed pertinent to the study (Stallings & Mohlman, 1988).

Researchers investigating teacher-student interactions and happenings in the classrooms on a large number of teachers favor direct observation (Galton in Keeves, 1988, p. 475). Four types of instruments used for collecting such data include category, sign, rating and checklists (Borich & Madden, 1977; Cohen, 1976; Medley & Mitzel, 1963; Rosenshine & Furst, 1973; Simon & Boyer, 1974). Elements that differ among these instruments are (1) their recording procedures, (2) their items or categories examined and (3) the type of instrument used to collect the data (Hook, 1981). Category instruments require that the data be recorded each time it occurs, using frequency counts. Sign instruments use some type of score or tally for behaviors that occur during specific intervals. Rating instruments require the observer to judge the behavior observed, usually at the end of the observation period. Checklists are often used when the behaviors to be recorded are known in advance (Hook, 1981).

Consisting of a list of statements about the behaviors to be examined, the checklist documents the absence and presence of the behaviors (Cartwright & Cartwright, 1974; Stalling & Mohlman, 1988). Its disadvantage is the inability to provide detailed descriptions of the context in which the behavior occurs. The checklist should be used when behaviors are known in advance. An advantage to using them is that the observer will less likely overlook an essential behavior (Cartwright & Cartwright, 1974). Although

the checklist provides no information on the frequency and/or quality characteristics of the behaviors being observed it does, however, answer the "what" question for the researcher (Stallings & Mohlman, 1988). With a checklist, the researcher can report what behaviors occurred and what behaviors did not occur during the observed period.

Rating scales offer the same information as the checklist but in further detail. They capture information on the frequency and/or quality characteristics exhibited during the observation (Cartwright & Cartwright, 1974; Stallings & Mohlman, 1988). Observers rate the behaviors on a scale "indicating the degree to which the behavior is present" (Cartwright & Cartwright, 1974, p. 100). In that observers are to rate these behaviors, judgments as such introduce the possibility of error. Thus, "the major difference between a checklist and a rating scale is that the observer is merely indicating presence or absence of a behavior with a checklist, and he is indicating his judgment about the frequency and/or quality characteristics of the performance when he uses a rating scale" (Cartwright & Cartwright, 1974, p. 100).

Gage (1972) suggested that researchers focus on specific aspects of a teacher's behavior rather than trying to focus on entire process of teaching at once. He emphasized the importance of studying the specific skills of teaching for their importance in developing "tools of the trade" for teachers to improve their instruction. He saw the objective of finding these tools to be effective in making teaching more manageable and to inform those that might be interested in entering the teaching profession (Rosenshine & Furst, 1973). Unfortunately, the problem with such research, as is apparent today, is that the development of such "tools of the trade" is seldom completed.

> Just as it is relatively easy to develop new observational systems, it has been fairly easy for educators to develop lists of teaching skills. Unfortunately the teaching skills, like the observational systems, are seldom validated against measure of student growth. Yet there is enormous potential value in research on tools of the trade. As a result of the operationally oriented approach to teacher training within the last decade, there exists a large number of teachers who have already received training in a variety of skills. Fortified with acceptable criterion measures, investigators could use existing observational systems to study the behaviors of these teachers and relate the skill-relevant behaviors to the measures of growth, and they could also compare the behavior and the outcomes for trained and untrained groups of teachers (Rosenshine & Furst, 1973, p. 126).

By the 1970s, the observation process introduced by Ned Flanders was being used widely. According to Rosenshine & Furst (1973), most of the earlier classroom observational instruments did not do much more than document the behaviors observed in the classroom. Studies during the 1990s brought attention to multidimensional instruments and various instruments capable of tracking changes in behavior and instruction. More current observation techniques represent a wide variety of styles (Galton, 1988). Today, classroom observational instruments are abundant with many sharing design and qualities of others so much that the "distinctions among types of instruments become blurred" (Rosenshine & Furst, 1973, p. 132).

Several studies examined the relationship between various classroom activities and measures of student growth. Of these studies, eleven are presented in *Mirrors for Behavior III* (Simon & Boyer, 1974). Three of the studies observed small numbers of classrooms yielding inadequate statistical analysis. Several examples from Simon & Boyer's *Mirrors for Behavior III* (1974) of systems used to observe classroom settings will be examined as well as several systems created since the review by Simon & Boyer. Current observation systems also will be examined as they introduce multidimensional

and tracking or trend capabilities. The following section will review and discuss the variety of these studies and various observational systems.

One of the earliest classroom observation systems, the *Social-Emotional Climate Index* was created by Withall (1949). Modified from the earlier pioneering work by Anderson & Hank (1969), known for his comprehensive system for examining the effects of teacher behaviors on student behaviors, Withall's system was used to provide feedback to teachers and their supervisors for improving the social-emotional climate of the classroom. The system was designed only to capture teacher behaviors that affect the climate of the classroom. Teacher behaviors observed and coded included: (a) learner supportive, (b) problem-structuring, (c) neutral, (d) directive and (e) disapproving and (f) teacher, or self-supportive statements or questions (Simon & Boyer, 1974).

Testing his instrument, Withall (1949) included a total of 23 seven-minute excerpts and five full class lessons. Using his protocol of teacher-statements (learner-centered, teacher-centered or neutral), patterns were obtained for each of the teachers reviewed and researchers attempted to make judgments to interpret these patterns. For example, one teacher appears to use learner-centered statements much more than the others of the recorded period. Researchers interpreted this teacher, through recorded excerpts, to be one that offers verbal support and praise to the students. This teacher had the lowest proportion of directive, reproving and self-supportive statements. Through analysis of all teachers recorded, researchers infer that her method for facilitating learning would probably include (a) keeping learners aware of the objective of the lesson and (b) maintaining a positive and "helpful" attitude with the students. This interpretation

was made from the review and analysis of the comparison of the teachers reviewed (Withall, 1949).

From Withall's (1949) findings upon testing his instrument, he concluded that it was possible to (a) assess and describe classroom climate; (b) train individuals to use the criteria and obtain adequate measure of agreement on the categorizing statements; (c) obtain, through the categorization of teacher-statements, a valid measure of the social-emotional climate of a group; (d) identify different patterns of verbal behavior used by several teachers; (e) assert that "statements categorized by the climate index as likely to produce 'positive' feelings tend to be similarly categorized by impartial observers and tend to be reacted to with 'positive' feelings by the individual to whom they are addressed" and vice versa for 'negative feelings' (Withall, 1949, p.360).

Developed in the 1950s, the *Instructor Observation Checklist* by Joseph Morsh was one of the earliest systems that examined the relationship between student achievement and teacher behaviors. Although not specifically designed for the traditional classroom experience, the three checklists, to be simultaneously coded by three separate observers, were intended to capture "instructor cognitive behaviors [how well the instructor 'defines terms,' 'explains,' 'asks questions of students,' etc.], instructor nonverbal behaviors and activities [categories such as 'stands behind desk', 'stands at board', 'demonstrates', 'uses the board', etc.] and student nonverbal behaviors and activities [categories such as 'student raises hand,' 'talks, 'answers or asks questions,' 'ignores instructor, 'yawns,' or 'smiles'" (Simon & Boyer, 1974, p. 425).

Ryan's *Teacher Characteristics Schedule* was a self-report survey that consisted of 300 multiple-choice and checklist items on various behaviors related to "personal

preferences, self-judgments, activities frequently engaged in, biographical data, etc" (Biddle & Elena, 1964, p. 79). Three booklets for the teacher characteristics schedule were developed—one for elementary teachers, one for English-social studies teachers, and one for math-science teachers. Ryan's dimensions of observed teacher behavior included "harsh-kind, aloof-responsive, stereotyped-original and evading-responsible" (Biddle & Ellena, 1964, p. 74).

The *Teacher Behavior Observation System* and *Student Behavior Observation System,* designed by Perkins (1964), were to be used as "companion instruments" by two observers simultaneously to collect data in two-minute samples (Simon & Boyer, 1974). The teacher instrument collected data with categories similar to those of Flanders System of Interaction Analysis along with various teacher roles within the classroom setting. The student instrument captured, in the two-minute sample, information about student activities and the "student learning environment (…discussion groups, class recitation, individual seat work and so forth)" (Simon & Boyer, 1974, p. 477).

Observational systems progressively became more complex. Spaulding (1967) designed three observational systems with two to be used as companion instruments. The *Coping Analysis Schedule for Educational Settings* (CASES) and the *Spaulding Teacher Activity Rating Schedule* (STARS) were designed to focus on individual students in a point-time sample. The systems have been useful in the selection of behavior modification programs for children as young as two years old. The third instrument by Spaulding, the most applicable to classroom settings, was the *Transaction Sample: Classroom* (TRC). This instrument was designed to "discover affective classroom correlates of pupil self-concept, academic achievement and creative thinking in

elementary school classes" (Simon & Boyer, 1974, p. 595). With 144 combinations of 90 classifications, the system provided a variety of possible behaviors captured and analyzed with a required eight hours of recording for every hour of classroom interaction (Simon & Boyer, 1974).

The *Step Observation Schedule* (STEPOS) created by Wallen, Moohr, Hall and Weisberg (1969) was also a time-sampling instrument. Ten major categories capture information on the verbal behavior of teachers and students (Simon & Boyer, 1974). The instrument used a rating system that codes and weights behaviors depending on the amount of time and the number of students involved in the particular behavior being observed. The instrument was to be used as part of a battery of tests that include "personality measures, attitudinal measures, supervisory ratings and projective tests designed to determine relationships between attitudes, behavior and achievement of black and white students and teacher behavior" (Simon & Boyer, 1974, p. 643).

Denny, Rusch & Ives' (1969) *Classroom Creativity Observation Schedule* (CCOS) was designed to examine the relationship between teacher behavior and student creativity. The dimensions reviewed include "climate (motivational climate, pupil-pupil relationships, teacher-pupil relationships and pupil interests); teaching-learning structure (including pupil initiative, materials available, adaptation to individual differences); and specific structuring (including encouragement of divergent thinking, unusual responses and uniqueness)" (Simon & Boyer, 1974, p. 219).

The *Flanders System of Interaction Analysis* (Flanders, 1970) is probably the most widely used classroom observation system. The system, though containing only ten categories has been praised for its proven usefulness in research and in teacher training. It

servers as a way to gather information to guide conversations with teachers regarding their teaching behaviors and the effects of those behaviors on the students they teach. The system has seen revisions by Flanders as well as by many other researchers (Simon & Boyer, 1974).

The *Flanders System of Interaction Analysis* and other studies reviewed by Simon & Boyer (1974) have covered various methods of observational systems including a category system to review measure of social emotional classroom climate (Withall, 1960); companion instruments examining the behavior of both student and teacher (Perkins, 1964; Wallen, Moohr, Hall & Weisberg, 1969); an examination into the relationship between teacher behavior and creativity (Denny, Rusch & Ives, 1969); and an instrument examining conversations with teachers regarding their teaching behavior and the effects of those behaviors on the students they teach (Flanders, 1970).

Research incorporating various observational systems since the mid-1970s and the Simon and Boyer (1974) review include studies by Ball, Camburn, Correnti, Phelps & Wallace, 1999; Brophy & Evertson, 1976; Coker & Coker, 1979a, 1979b; Kennedy, Ball & McDiarmid, 1993; Medley, 1979; Soar and Soar, 1982; Stallings, 1977; Stapleton, LeFloch, Bacevich, & Ketchie, 2004; Valli, Raths, Rennert-Ariev, 2001; and Weiss, Pasley, Smith, Banilower & Heck, 2003.

Brophy and Evertson's *Texas Teacher Effectiveness Project* (1976) used several instruments for collecting both high and low inference variables. An earlier system by Brophy and Good (1974) was modified and used as a low inference coding system. In this system sequences of teacher questions, student responses and teacher feedback were coded. These included quality, types of and reactions to various responses and questions.

Another instrument was created to capture variables that occurred during structured lessons. It included how the lesson was introduced, how material was presented, feedback and opportunity for practice. Also included in the Brophy-Good system was a rating scale developed by Emmer and Peck in 1973. This instrument consisted of a set of 12 five-point rating scales on classroom interaction. These scales included variables such as "positive/negative affect, clarity, enthusiasm, use of student ideas, types of questions asked, etc." (Brophy & Evertson, 1976, p. 184).

Along with several other rating and checklists, Brophy and Evertson (1976) interviewed and administered several questionnaires to measure presage variables. Items that were appropriate for the checklists and rating scales were presented on the questionnaires while those that were more open-ended were included in the interview. The researchers approached this study "with the intent to measure anything that seemed likely to correlate with student learning gains;" thus, the inclusion of multiple instruments and measures (Brophy & Evertson, 1976, p. 174).

A study out of the Stanford Research Institute utilized a protocol called the *Classroom Observation Instrument* (Stallings, 1977). The purpose of the research was to investigate the relationship between child outcomes and classroom instructional practices. Three instruments were used including a physical environment and classroom checklist and a five-minute interaction form. The form was built on the Flanders Interaction Analysis model (1970) which was designed to record interactions between the teacher and the student (Cartwright & Cartwright, 1974).

"The *Observation Schedule and Record, Form 5, Verbal* (OScAR 5V) is a verbal category system designed to describe the classroom learning environment according to

the relevant frequencies of 80 different kinds of events in classroom interaction" (Medley, Coker & Soar, 1984, p. 245). "OScAR records may be scored on eight keys which were empirically derived by factor analysis" (Medley, Coker & Soar, 1984, p. 254). The eight keys include "(1) managing behavior, (2) rebuking behavior, (3) permissive behavior; (4) listening behavior; (5) lecturing behavior; (6) question source; (7) question difficulty;(8) question quality" (Medley, Coker & Soar, 1984, p. 255-56). An elaborate system coding eighteen categories that yield frequencies on 75 types of events is used for coding behaviors.

The *Classroom Observation Keyed for Effectiveness Research* (COKER) is used to measure low-inference items of teacher/student behavior in the classroom (Medley, Coker & Soar, 1984). The items on COKER have been reduced and adapted from more than 1344 category and sign items on five separate observations instruments to 18 "teaching effectiveness competency statements" (Gordon & Yocke, 1999, p. 47).

The ideal procedure for administering the COKER includes advance scheduling and early arrival early for the observer to become oriented to the situation. Once information on the data sheet has been completed and the observer has oriented himself to the classroom, the 5-minute observation period should begin. The observer codes information from this observation period in Section A. Then, Section B is to be completed from memory, following the 5-minute observation. At this point, one observation has been completed. This scenario is completed a second time, bringing the total observation time to 10 minutes per visit, although as much as 20-25 minutes may be spent in the classroom (Medley, Coker & Soar, 1984). Findings from the analysis of the COKER yield various measures of a teacher's classroom (Medley, Coker & Soar, 1984).

"*The Climate and Control System (*CCS) is a sign system that includes items relating to the nature of classroom structure, teacher and pupil control strategies, and teacher and pupil affective behaviors, both positive and negative" (Soar and Soar, 1982 in Medley, Coker & Soar, 1984, p. 245). "The teacher behaviors scale is intended to represent increasing degrees of coerciveness on the teacher's part-both verbal and nonverbal" (Medley, Coker & Soar, 1984, p. 256). Other information collected includes the attention the teacher gives to students, the types of groupings observed, the interest and involvement in the lesson or activity, and types of student behaviors recorded as well as other interactions. A separate page of the CCS is used to code teacher affect expressions. The extensive instrument collects a wide variety of measures in examining the affective behaviors of teachers and students as they relate to classroom structure (Medley, Coker & Soar, 1984).

The *New Tools for Research on Instruction and Instruction Policy*, a web-based teacher log, was created from the work of Michigan State researchers during the 1990s. Authors of this working paper by Ball, Camburn, Correnti, Phelps & Wallace (1999) report on the development of the pilot testing of their instrument designed to collect data on daily instruction from teacher logs. Included in the pilot were seven teachers in two schools who reported on 29 lessons in mathematics and reading. Twenty-four of these lessons included observed and detailed narrative descriptions completed by project researchers**.** According to the report, results of the teacher log revealed "insight into: (a) the work of particular students and student groups; (b) the nature of student activity; (c) the nature of teacher activity; and (d) the topics and materials used in instruction" (Ball, et. al., 1999, p. 16). Researchers were cautious about capturing the more inferential items

such as the content and how it was presented to students, what the students and teachers were doing during the lesson, the purpose and student engagement through teacher self-reports on the log. The authors suggested that more work needed to be done in these areas before presenting them for use in a web-based instrument (Ball, et al., 1999).

Valli, Raths, and Rennert-Ariev (2001) report on a survey they designed to "(a) gather information from beginning teachers about their preservice and induction learning experiences; (b) develop theoretically and empirically based constructs of teacher preparation, induction, knowledge, beliefs, and practice; and (c) determine relationships among three sets of variables (teacher preparation, teacher knowledge and practice, and student learning)" (Valli, et al., 2001, p. 1). The survey was administered to beginning teachers in Tennessee and Connecticut. Approximately six hundred 1st, 2nd and 3rd year teachers teaching grades 3-8 in reading and/or mathematics responded.

A preliminary finding of Valli and colleagues (2001) reports that the estimates of teacher effectiveness among beginning teachers is more varied among math instruction than other subjects. This finding became cause for further analysis to center on effectiveness of teaching in the area of mathematics. Measures included an efficacy scale, an adaptability/flexibility scale, a math preparation scale, and mentor frequency examining various dimensions of teaching. The authors conclude: "…basic findings indicate that beginning teachers are more successful in their teaching of mathematics if they (a) believe teachers can have an impact on student learning, (b) help students make sense out of mathematics, (c) were taught how to do this in their teacher preparation program, and (d) received continued support in their first years of teaching by an experienced mentor" (Valli, et al, 2001, p. 6).

The major purpose of another study, the *Study of K-12 Mathematics and Science Education in the United States: Inside the Classroom* study (Weiss, Pasley, Smith, Banilower & Heck, 2003), was to provide a current snapshot of what was happening in math and science classrooms across the country. The instrument used to capture this information was adapted from a classroom observation instrument developed by Horizon Research, Inc. as part of an evaluation instrument for the National Science Foundation's Local Systemic Change initiative. Originally the instrument was designed to "assess the quality of the design and implementation of mathematics and science lesson" (Weiss, Pasley, Smith, Banilower & Heck, 2003, p. ix). Modifications to the instrument included an interview protocol to gather more data on what factors were involved in the planning phases of instruction (Weiss, et al., 2003).

The *Inside the Classroom* study followed a selected sample of 31 middle schools. The analysis is based on a sample of 364 mathematics and science lessons from grades K-12. The study's rating scale ranged from ineffective instruction to exemplary instruction, using observer ratings categorized from low, to medium and high quality. Of the math and science lessons observed, 15% were estimated of high quality, 27% were medium quality and 59% were low quality (Weiss, et al., 2003).

The protocol produced overall ratings for four lesson components: design, implementation, content addressed, and classroom culture. Key factors (captured through rating scales of various components) distinguishing high quality lessons from low quality lessons were reported in their ability to "(a) engage students with the mathematics/science content; (b) create an environment conducive to learning; (c) ensure that all students have access to the lesson; and (d) help students make sense of the

mathematics/science content" (Weis, et al., 2003, p. *xi*). Quantitative and qualitative analyses were employed. Field notes and interviews were themed and analyzed by the researchers. Rating scales from the classroom observations and teacher interviews were weighted and analyzed quantitatively (Weiss, et al., 2003).

The authors conclude that based on their findings, "…the nation is far from the ideal of providing high quality mathematics and science education for all students…both the lesson snapshots and teacher reports on what influence their lesson designs, have implications for the preparation and continuing education of the mathematics/science teaching force, and for the support provided to teachers" (Weis, et al., 2003, p. 104). They conclude that teachers need (a) a vision for effective instruction to guide their lessons; (b) support materials for better targeted assistance; (c) workshops and professional development reflecting high quality instruction; (d) equity in quality of instruction for all students; and (e) better alignment of preparation, curriculum, student assessment, profession development and teacher evaluation policies (Weis, et al., 2003).

Stapleton, LeFloch, Bacevich, & Ketchie (2004) designed a classroom observation instrument specifically to evaluate two curriculum packages. To adequately evaluate the programs, the research team created a set of data collection instruments. Data collected for analysis included responses from an interview of administrators and teachers, student focus groups, samples of student work, and student achievement scores. Sixty-six math and English classes, grades six through ten were included in the study. Forty-one were treatment classes, with the remaining 25 serving as control classes. Observations included a pre-observation interview for information on the lesson to be observed; the observation; and a post-observation interview (anything the teacher wanted

to elaborate on that occurred during the lesson). The observation instrument included a section to capture demographic information and a narrative log, designed with a specific structured coding system for activities, engagement or actions occurring, instructional strategies, materials used and behaviors observed (Stapleton, et al., 2004).

Findings of Stapleton and colleagues' (2004) research reveal that a large number of teachers from the experimental group included more instructional activities into their classes and the nature of the activities were different between the control and experimental groups. Also, the treatment classes reported a "much higher level of student engagement, teacher time devoted to instruction and a lower level of disruptive student behavior" (Stapleton, et al., 2004, p. 22). "The benefits of this observation form stem from the desire to produce easily assimilated and quantifiable output from the observations, while still addressing the complexities of the classroom experience" (Stapleton, et al., 2004, p. 22).

Summary and support for observational systems

All classroom experiences cannot be observed. In addressing the complexities of the classroom experience, other measures may be necessary to ensure an accurate picture of what is happening during a lesson. To aide in creating an accurate portrait of the classroom, the perceptions of teachers should be considered. Philosophical beliefs of teachers may differ, but among these philosophical differences a teacher's beliefs are the "cornerstone of [his] teaching practices and beliefs concerning teaching and learning" (Campbell, Kyriakides, Muijs & Robinson, 2003, p. 52). These beliefs also affect how instruction is presented. Campbell, et al. (2003) write, "Teachers' beliefs may develop

into a coherent philosophical system that directly influences their overall classroom behavior. A teacher's own philosophy is thought to function as a filter influencing decisions and actions made before, during and after instruction" (p. 52). Biddle & Ellena (1964) suggest that **"**the teacher views her behavior as a product of interaction between situation demands and personal factors (such as educational philosophy, needs, beliefs, values and motivations)" (p. 11). Given that the philosophical aspects of the teacher and beliefs held within the individual affect the type of instruction delivered, the interactions that occur and the process of daily functions in the classroom, finding a method for capturing this data is imperative when considering aspects of effective and ineffective teacher behaviors and characteristics.

Observational and survey instruments or a combination of "companion instruments" are designed to capture various dimensions of more than one individual at a time. The range of classroom activities as well as the measures of student growth varies from study to study, but all provide a methodical approach to objectively capturing information on behavioral aspects that were complex and often difficult to measure. The methods used to obtain these findings included rating scales, category systems using both frequency and duration and historical documents. The most significant of the findings were determined to have been obtained using the rating scales, although not all items on either the rating scales or the category instruments yielded consistent results (Rosenshine & Furst, 1973).

> Even if one's approach to observation is grounded in abundant and well-honed theory and research, the problems of instrumentation are sufficiently complex that it cannot be assumed that the items, scales, and formats chosen for an instrument will be the most functional ones for the situation. …The optimal strategy at this point would be to use a variety of instruments in every study (Rosenshine & Furst, 1973, p. 136).

Foster's (1996) response, similar to Rosenshine & Furst, suggests that often the most appropriate method for a study might be a combination of different methods. He further remarks that many times a combination of quantitative and qualitative methods may be appropriate (Foster, 1996). Gage's (1978) remarks on quantitative versus qualitative methods:

> It seems unlikely that either the qualitative or the quantitative research [worker] will accept the other as a replacement. …it is important to note that the two kinds of research operate in different contexts—the context of discovery and the context of justification. The qualitative researcher can discover new phenomena and relationships or create new hypotheses. The quantitative researcher is better able to test, validate, or justify the hypotheses (p. 83).

Systematic observation in classrooms involves creating categories of classroom behaviors and classifications to be listed on an observations schedule (Foster, 1996). An example of such an instrument is the Flanders Interaction Analysis Categories (Flanders, 1970). An important point to make is that these classifications and categories are created prior to data collection. The techniques for collecting the data may vary, but all involve a preset, standardized observation form to capture the data (Foster, 1996).

> Quantitative approaches to observational research in schools have aimed to describe in numerical terms some of the key pattern and regularities of school life. Researchers adopting this approach try to produce accurate quantitative data on the frequency, duration, intensity, and sometimes the quality of particular behaviors or patterns of interaction occurring in schools (Foster, 1996, p. 3).

This review of observational systems provides information on the historical context of using observational systems within classrooms, specific to observing behaviors as they relate to varying outcomes. Of the literature reviewed no one system stands out as most effective in examining teacher behaviors and their relationships with student

outcomes. Many examples provided promising evidence of the possibility of designing accurate tools to measure and assess teacher effectiveness in the classroom setting This summary has created the need, as suggested by Rosenshine & Furst (1973), Gage (1978) and Foster (1996), to use more than one observational system in an attempt to accurately capture data when reviewing classroom practices of the teacher as they relate to student achievement outcomes.

As mentioned in Rosenshine & Furst (1973), "Current observational instruments disregard materials being read, the assignments students write, the teacher's use of written or oral material, and the physical features of the room, such as seating arrangements and lighting. These additional classroom events and characteristics might be profitably incorporated as variables in research on student growth" (pp. 165-66). This consideration also was suggested by Hook (1981). He suggested, (regarding using Flanders Interaction Analysis Categories -FIAC), "When reviewing the verbal interaction in the room….it might be necessary to consider the way the room is organized. Other behaviors that can be supported through classroom drawings might be the teacher-centered/student-centered discussions (by way of the arrangement of desks), etc." (Hook, 1981, p. 89). To address concerns presented by Rosenshine & Furst (1973), Gage (1978), Hook (1981), Foster (1996) and others, the proposed study will employ methods utilizing both qualitative and quantitative approaches.

Confronting methodological challenges

How best to determine effective teaching practices presents a plethora of shortcomings and complexities, most of which are methodological challenges. According

to Millman (1997), psychometric problems with the high-stakes tests as well as the reliability of the measurements used to determine teacher effectiveness should be causes for concern. Other shortcomings include the generalizability of many of the measures, the accuracy of the measures when weighed against the curriculum being taught, weak research designs and weak measures to capture the data (Cruickshank, 1990). These challenges remain but are much more manageable than they were fifty to sixty years ago. Pioneering researchers have grappled with and solved many of the issues that once were major stumbling blocks.

Historically, the ability to systematically gauge student learning, compare that learning from year to year and link that learning to a specific teacher has been lacking. To accurately gauge student learning, one first needs to have a database that effectively stores test data with unique student identifiers. The test also must be aligned with the curriculum taught. These historical impediments are now just that, historical. Computers provide adequate and ample databases to house the information. Today, curriculum specialists and test makers consult and are capable of creating tests that capture the curriculum being taught. With the mandate of No Child Left Behind, students are tested annually. Suddenly, the ability to gauge student learning through gains calculated from annually administered standardized test and link that learning to the teacher has become feasible.

Millman (1997) suggests four contemporary approaches to examining teacher effectiveness through methods of student learning gains. These approaches are the Teacher Work Sample Methodology, the Dallas Value-Added Accountability System, the Kentucky Instructional Results Improvement System and the Tennessee Value-Added

Assessment System. These four approaches confront many of the challenges presented in previous research such as the context within the classroom, the unreliability of measures used to define teacher effectiveness and possible alternative options to the standardized tests (Millman, 1997).

The Teacher Work Sample Methodology (TWSM) originates from Western Oregon State College. This methodology was adopted by the State College in 1988 to help evaluate its teacher preparation and licensure programs. It has since been refined for use as a means of examining the ongoing research programs on teacher effectiveness. The TWSM researchers argue that the use of standardized tests as a true measure of student learning, while ignoring the contextual issues under which teaching and learning occurs, is a grave mistake. For these reasons, they consider "attempting to connect student learning to the work of individual teachers [is] indefensible" (Schalock, Schalock & Girod, 1997, p. 16).

There are four procedural undertakings of the TWSM: (a) pupil assessment is linked specifically to the outcomes demonstrated/attempted by the teacher, (b) it uses criterion instead of norm referenced measures, (c) gains are calculated on a student-by-student basis allowing for differences among high- and low-scoring pupils, and (d) classroom and community contexts are included in the measurement. Some concerns about this approach involve validity and reliability of the measures used as well as the use of the data for various purposes including the support for teacher licensure issues, examining teacher effectiveness and the evaluation tool for improving teacher preparation programs. It would be highly unlikely that an instrument that considers teacher licensure would be able to provide the same measures to examine teacher effectiveness or teacher

preparation programs. The three are not the same. Using a single method for these three purposes is not a sound methodological practice. Despite these concerns, Airasian (1997) asserts that the benefits to teacher education programs as well as continued professional development programs support the justification for the use of TWSM. He also suggests a move to improve and extend the methodology to encompass all ranges of teaching, instead of only beginning teachers.

The TWSM approach deals with the contention of the importance of improving student achievement as a primary goal in examining and determining teacher effectiveness. The approach is theoretically sound and has a practical focus but lacks certainty in dealing with larger issues such as policy implications and for guiding future research (Stufflebeam, 1997a). Stufflebeam (1997a) purports TWSM as a "promising but limited technique for use in teacher education and formative evaluation of teachers" (p. 60). He also asserts that the approach is superior as a "systematic and useful means for assessing teacher effectiveness based on pupil outcome data" (Stufflebeam, 1997a, p. 61).

Millman (1997) commends Western Oregon for their focus on teacher improvement but their methodology suffers at the expense of their accountability efforts. In referring to the Oregon Works Sample methodology, Darling-Hammond **(**1998) commends the group for looking at teaching within the context of teachers' goals, the classroom and student learning. A weakness of the Oregon Works Sample methodology is the high variability of the quality of assessments devised by the teachers. Darling-Hammond states, "As it evolves, it could be a valuable tool for preparing and assessing both beginning and veteran teachers" (p. 257).

A second approach presented by Millman (1997) is the Dallas Value-Added Accountability System (DVAAS). This approach was heralded as a fair and equitable value-added accountability system by the Dallas Public Schools during the early 1990s. The model used multiple regression to construct longitudinal growth curves of its students on the state's norm referenced tests. The Board of Education recommended the model to be extended to include other variables in addition to the norm reference tests and also to measure teacher effectiveness. The model was revised and presently uses a combination of multiple regression with hierarchical linear modeling to address the Boards' recommendations (Webster & Mendro, 1997).

In the current model, DVAAS analyzes student outcome variables in a two-stage process. First, multiple regression controls the effects of the ethnicity, gender, language proficiency and socioeconomic status called "fairness variables" (Webster & Mendro, 1997, p. 82). In the second stage, the two-level HLM controls the effects of prior student achievement or attendance and the influence of any aggregated school variables. Issues surrounding this approach include the consideration of a three-level HLM analysis instead of the two-level HLM analysis employed when using it to measure teacher effectiveness indexes. Another issue is how the Dallas model deals with missing data. By removing students with incomplete data, sample size becomes an issue.

Thum and Bryk (1997) reviewed the Dallas model and concluded that the work of Webster and Mendro (1997) included the necessary sophistication to begin to wrestle with the many issues surrounding such a model, but should be approached with great caution. In discussion of the technical issues surrounding the validity of the Dallas model they concluded, "It is easy to do this badly but very hard to execute such a program well.

As much as Webster and Mendro have accomplished, we are still not there" (Thum & Bryk, 1997, p. 108). Sykes (1997) also presented his concerns of the Dallas model from a different angle. He addressed the never-ending issue of whether a standardized test is the correct instrument to measure student learning. Webster and Mendro's DVAAS model, still in use, has taken these critiques seriously and continued to improve upon their model.

The third approach to examining teacher effectiveness through methods of student learning gains was prompted by the Kentucky Supreme Courts' order to establish an equitable and efficient system of education. Advanced Systems for Measurement in Education won the request for proposals in 1991 and was awarded a 5-year contract to develop, implement and administer a new system to respond to the order. The primary goal was to "motivate educators and the public to dramatically improve student learning" (Kingston & Reidy, 1997, p. 191). This accountability undertaking addresses the needs of the Kentucky education system, which includes holding the schools accountable for the success of all students. This school-based accountability, known as the Kentucky Instructional Results Information System (KIRIS), is measured against baseline data such as attendance, retention and academic achievement relative to that particular school. Incentives are provided for schools exceeding the goals, and assistance is provided to those not reaching them.

Stufflebeam (1997b) refers to efforts made by Kentucky as the "forefront of modern trends in assessment" (p. 219). He reports of Kentucky's efforts to shift assessment of student outcomes from multiple-choice tests to performance-based practices. The impact of these efforts on teachers, schools and the state education system

is fraught with problems. Teachers expected timely feedback to help immediately in their classroom activities but this was not part of the design. School leaders express concerns about whether the incentives and sanctions are fair and appropriate or consider contextual issues within each school community. These issues along with the retraction of incorrect indexes have caused embarrassment to the state. The state has been faulted with "putting the airplane together while trying to fly it" by one educational official (Stufflebeam, 1997b, p. 223). The initiative was heralded for its ambitious attempts to recreate the education system through performance-based assessment. The work deserves recognition but, riddled with deficiencies and mistakes, states should not move carelessly or haphazardly with unsubstantiated assessment systems to solve their educational problems.

The fourth approach to be discussed as a model of evaluating teacher effectiveness was adopted in the state of Tennessee. The Tennessee Value-Added Assessment System (TVAAS) was the first of its kind to be adopted as a statewide assessment system (Ceperley & Reel, 1997). The main purpose of this model was to provide summative information on the effectiveness of a school, school system or teacher in adequately moving students forward academically (Sanders, Saxton & Horn, 1997). The TVAAS model uses a mixed-model methodology to estimate effects of schools, the school systems and individual teachers on the academic growth of students (Ceperley & Reel, 1997). This approach is possible given the advantage of the longitudinal nature of the data (Sanders, Saxton & Horn, 1997).

Issues of sample size that plagued the Dallas model are not a problem with the TVAAS model. Consider a teacher assessing her class. The more scores that are available

on the student, the better and more accurate the assessment (reducing the error) of that individual student. The same approach can be considered when examining estimates for teachers. Solomon, White, Cohen & Woo (2007) provide a great explanation of this concept:

> If a teacher has only 10 students for whom she has reliable test data (resulting in a larger standard error), her final effect estimate will give more weight to the teacher average than to her individual point estimate, moving her final estimate more towards the average, even though her individual point estimate may be much higher or lower than the average. If a teacher has 30 students with reliable data, her effect estimate will give more weight to her individual point estimate, pulling her estimate away from the average and more towards her individual score. If a teacher with a larger and more reliable set of data (which is reflected in a smaller standard error) has an individual point estimate that is detectably different from the average teacher, her overall estimate will reflect this difference. Because all teachers are considered to be equal in their effectiveness (i.e., at the average level) until the quality and quantity of the data for their classrooms (historical and current year) pull their estimates away from the average, teachers categorized in the tails of the distribution have the most accurate estimates and relatively large effect sizes compared to their standard errors. In other words, all teachers who have scores that are above or below the average have them because they had relatively large differences between the scores of their students and the mean score, and enough reliable data (small standard errors) to pull them away from the average. Otherwise, teachers are determined to be closer to the average because of less reliable data, or small differences from the mean, or both (p. 9).

This process is known as “shrinkage” and these estimates allow researchers to compare the relative magnitude that teachers have on the achievement of their students. The TVAAS mixed-model approach allows the inclusion of all available data, even if incomplete. The model is “designed to provide rigorous protection against the severe misclassification of a school’s, system’s or teacher’s influence on student gain” (Sanders, Saxton & Horn, 1997, p. 143). Teachers are assumed to be average among other teachers

within the district until the data pulls the teacher in one direction or another away from that average.

The TVAAS model asserts the following statistical claims:

> (a) students serve as statistical blocks; (b) all data that is available is used and imputation is not required; (c) longitudinal analysis improves the efficiency of the model from year to year; (d) repeated measures across subjects provides similar benefits; (e) gains can be estimated from a model that uses scores, not gains; (f) providing shrinkage estimates of teacher effects forgoes misclassification of individual teachers; and (g) the layering model improves the efficiency of the estimate of teacher effects (Sanders, Saxton & Horn, 1997, p. 144).

These statistical claims for use of the model "provide linear metrics, are strongly related to defined curricular objectives, and possess appropriate sensitivities" (Sanders, Saxton & Horn, 1997, p. 161). The model has been proven to be flexible and can accommodate different assessment instruments including both norm-referenced and criterion-referenced tests.

Darlington (1997) counters the TVAAS model's use of the mixed-model approach that suggests that 'less is more.' The TVAAS model only uses student achievement data, opting not to use data such as IQ scores, attendance and other personal factors of the student. Others that would defend a mixed-model approach would suggest that other factors considered to be constant across a child's school career should not be included in the model. Sanders, Saxton & Horn (1997) respond to Darlington's conclusions by reporting that the TVAAS estimates, when discussing the simplicity or complexity of various data sets, are capable of protecting against any "spurious misclassification by incorporating all available data into its multivariate, longitudinal model" (Millman, 1997, p. 179).

Walberg and Paik (1997) further conclude that the TVAAS model, although exemplary in its longitudinal, statistical approach lacks the capacity to improve education in that they offer no policy or compensation plans for the individual teacher's merit. In response, Sanders, Saxton and Horn (1997) assert that the TVAAS model presents the data that can then be reviewed and applied as needed by different school systems and states as applicable to a specific classroom, school, district or state. They further explain that within the state of Tennessee, TVAAS is only a part, although an essential part, of the educational assessment program in Tennessee and that the information the model provides cannot be obtained from other assessment instruments. Sanders, Saxton, and Horn (1997) conclude with the following remarks:

> It is our hope that as educational assessment and policy theorists begin to delve more deeply into this admittedly new approach to use of scaled achievement data, they will discover, as did our reviewers, that TVAAS is a methodology that is statistically rigorous, fair, reliable, and valid on which decisions about educational practice can be firmly based (p. 181).

Darling-Hammond (1997) suggests that the appropriateness of the four approaches could be answered with two essential questions: (a) Are the systems measuring the quality of schooling or teaching [or something else]? and (b) What are the effects and do they improve or harm teaching and learning? She echoes Millman's sentiment that the idea of improvement is good, but when attached to a model of accountability, turns bad. The Dallas Model did not meet the needs of either an appropriate evaluation or an improvement program. She listed concerns such as the mismatch between the curriculum and the test measures and the fact that several subject areas were not being addressed at all.

Darling-Hammond (1997) also suggests that the Tennessee Model stands up as unobjectionable in its ability to handle missing data and to parcel out effects of different teachers but further explains that for personnel evaluations, the model would not be appropriate. She notes the Tennessee Model's longitudinal approach as well as its use of analytic tools that allow the use of existing data are major strengths of the system and "as a research tool the Tennessee Value-Added Accountability System (TVAAS) has clear value" (p. 254).

From these readings, the various approaches provide a range of options in creating new ways to examine teacher evaluations that address and lessen the methodological challenges, providing incentives and sanctions for effective teaching, suggesting better practices to use in teacher preparation programs and improving professional development of teacher practice.

Having carefully reviewed the four current models of evaluation methods using student achievement data to evaluate the effectiveness of a teacher, I have chosen to use the TVAAS model for the following reasons. First, the model has at its disposal an entire state of student records and supports its claims using this massive longitudinal database. To err on the side of caution, this model also requires three consecutive years of data for analysis. As a convenience factor, the district in which I plan to gather the data is within the state of Tennessee and scores are readily available for analysis. Most importantly, from the literature reviewed, the TVAAS model presents the best case as being the most accurate and reliable method for exploring the questions presented in this study of teacher effectiveness.

The state of Tennessee has collected longitudinal data on students for almost three decades. Having this massive database available on students allows researchers to track progress of an individual throughout his schooling career through varying teachers and to track teacher progress from class to class (i.e., year to year).

Being situated in a district within Tennessee that has a mix of urban, rural and suburban schools presents a suitable area for such a study to be conducted. This variation in demographic settings will allow researchers to consider any aspects that might surface regarding the different backgrounds of students. Having a working relationship with both the district and individual schools allows for researchers to gain access to teachers for study. Also, working with the district on other educational projects has provided the researcher an opportunity to become acquainted with the appropriate staff responsible for allowing such a research endeavor to happen. The easy access of available data from this district and the relationship with the necessary players to make such a study happen makes Hamilton County an ideal choice of locations for this study.

And last but not least, the issue of TVAAS being the most accurate and reliable value-added model available at this time also explains its use in this study. Although other models are available and are being used in various research studies, TVAAS is still known as the best model for evaluating teacher effectiveness.

Framing the study

As the literature suggests, many factors play a role in student achievement. The researcher recognizes that factors, other than those proposed in this study, may have an impact on a teacher's effectiveness. However, given the scope and resources of this

study, the researcher has chosen to consider (a) planning and preparation and (b) instructional practices for investigating of variation in these behaviors as they related to the effectiveness of the teachers.

The ultimate goal of effective planning and preparation and effective instructional practices is that they lead to increased student achievement. Increased student learning is defined, for the purpose of this study, as student achievement test gains.

In defining teacher effectiveness, several things should be considered. Background experiences that a prospective teacher brings to the university setting, training that happens in that setting and planning that occurs once the teacher is within the profession all affect how the teacher will respond to students in her classroom. These experiences will be examined in the model's planning and preparation domain. Also to be considered in this domain are the attitudes teachers have regarding their teaching, other colleagues and the school environment. These areas will be pursued in this model's planning and preparation domain. Just as the aspects of planning and preparation of teachers play a major role in the student achievement of students in their classroom, equally important are the actions within the classroom. Instructional processes within the classroom will include teacher and student actions, as well as interactions. These actions include strategies used by the teacher, student groupings for various activities, as well as assessment procedures that determine what learning has occurred. These areas will be examined in the model's instructional practice domain.

Planning and preparation processes and instructional processes are not mutually exclusive. Teachers possibly may reflect on a lesson that was implemented, revise as needed in the planning phase and readdress the issue to achieve the intended instructional

goals for the students. The interconnection of the two phases suggests the opportunity for this to occur. Assessment might be an example of an activity that would be placed in the interconnecting sector of the two domains.

The framework (see Figure 1.1) will guide the study through the data collection and analysis phases, as well as in the written dissemination of the findings. Data collected will be grouped under various subcategories within the two domains of planning and preparation and instructional practices. Findings will be reported as analyzed by determining how student gains are affected, by the behaviors and actions represented in the two domains, by teachers at varying levels of effectiveness as determined by their teacher effect scores.

CHAPTER III

RESEARCH DESIGN

Purpose of the study

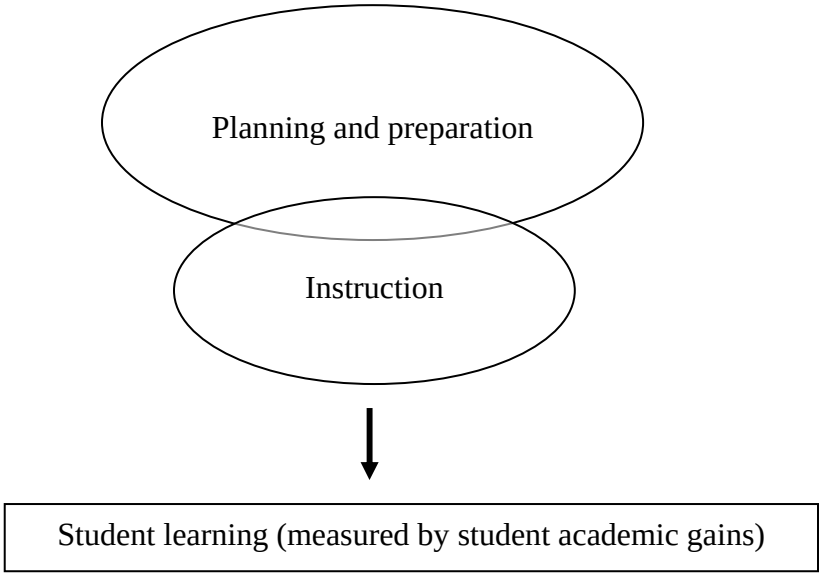
The purpose of this study was to investigate varying levels of effectiveness among math, reading and language arts teachers in grades 3-8, defined by their (a) preparation and planning and (b) instructional practices that lead to gains in student achievement.

Conceptual Framework

What aspects of teaching lead to higher student achievement? This study examined the domains of planning and preparation and instruction through surveys administered prior to and after a scheduled observation to shed light on this question.

Figure 1

The study's model of teacher effectiveness*



**Adapted from Hay McBer, 2000.*

Evidence of the two domains of study was captured through survey and observational data. Planning and preparation, for the purpose of this study, included the perceptions of the teacher towards teaching, instructional preparation and interactions with colleagues. Behaviors and practices exhibited in the preparation and planning domain also were collected and documented by trained observers during a scheduled observations. Aspects of this domain encompassed ideologies and experiences that were unique to the individual and might express concepts derived from various points in the teacher's life. Professional development also was considered a major function of the planning and preparation domain. Effective professional programs offer the teacher suggestions to better plan and prepare for effective instruction.

Observers reported on aspects of planning and preparation as they were present during an observed lesson. Evidence was collected that would support the stated objectives, the lesson's planning and organization and the activities and strategies used to provide evidence of an awareness of diversity and equity issues among students in the classroom. Comments from the reflective survey completed by the teacher also provided additional data for the planning and preparation domain essential to the lesson observed. Teacher reflections on the observed lesson provided insight to the observer on aspects of planning and preparation that might not have been observable behaviors.

In the instruction domain, researchers examined actions and behaviors that occurred during the lesson. Teacher actions included how the lesson was delivered, various strategies and materials used, groupings for instruction and other aspects of how the lesson was taught.. Student actions also played a role in how instruction was carried out. These actions included the discussions surrounding the lesson's topic, various materials used, student grouping and communication among teacher and students.

Other variables in the instruction domain included evidence within the observation such as the teacher's demonstration of content knowledge and a variety of strategies to reach diverse groups of students in a developmentally appropriate way. Content knowledge was examined in the summative observation instrument domain of planning. Behaviors that provided evidence that tasks and interactions within the classroom proved consistent with what was observed for the content examined were noted as such. Behaviors and strategies that inferred prior planning and thought presented by the teacher also were considered. The instruction domain closely examined these actions as well as the interactions observed and documented. The variety of strategies

designed to reach diverse groups of students included the use of various materials and activities to describe a single concept and working with small groups of students.

Interactions cross the domains. Often teachers reflected upon prior instruction and used that information to prepare for future instruction. (a) Planning and preparation and (b) instruction are not mutually exclusive. They cannot be separated into neat categories that are unrelated. Instructional behaviors are not isolated to a single incident. Thus, this study's framework provided for the interconnection of the two domains being examined. The model's goal was to identify areas within the two domains that affect positive student academic gain. Whether a behavior was found to be exclusive of a single domain or a behavior that blended within the two domains, it was captured and analyzed for its relationship to student gains.

There were other factors that played a role in student achievement that will not be addressed. The researcher is aware of these factors, and has decided to limit this study to the domains of (a) planning and preparation and (b) instruction. Other factors that related to positive student achievement will not be topics for this specific study.

Rationale for the use of a quantitative/qualitative approach

(a) Planning and preparation and (b) instruction are broad areas to define. Efforts to use a single test item to determine a teacher's perception of his teaching experience is virtually impossible. Several items were used to capture the intended meaning of the subcategories created under each domain examined. Subcategories of the planning and preparation domain included selected variables within classroom techniques and experiences, professional development, college training and experiences variables.

Instruction domain subcategories included other variables within classroom techniques and experiences variables and other variables relating to actions and behaviors occurring during instruction.

The design for this study involved factor analysis and multiple regression analysis. Through factor analysis, the many variables examined were reduced to a manageable number of interpretable factors for use in subsequent analyses. Regression analysis is designed to determine the relationship between the multiple variables derived from factor analysis and the dependent variable, the teacher effect score. With regression analysis, the factors created were referred to as predictor variables and the teacher effect. This study examined the relationship between these predictor variables and the teacher effect. Qualitative data collected was used to describe the participants of the study and served as reinforcing statements to support quantitative findings.

Research problem

Research question

The major research question addressed in this study was: What are the differences in (a) planning and preparation and (b) the implementation of instructional practices of teachers at varying levels of effectiveness? This question will be addressed using math, reading and/or language arts classes in grades three through eight in one Tennessee school district.

Participants and site selection

Criteria for eligibility for participation in the study included (a) the teacher must teach at least one of the three subjects being investigated (math, reading and/or language arts) and (b) he must have a teacher effect score. Participants were chosen from the Hamilton County school district in Tennessee, representing rural, suburban and urban schools in grades three through eight. They represented a purposive sample, chosen systematically from the group of teachers from the eligible participant pool. A total of two thousand nine hundred seven teachers were employed in Hamilton County during the 2001 academic school year. This number included K-12, music, art, physical education, vocational education, alternative achievement, gifted, band, library, modified resource, CDC and visually impaired teachers. The total number of teachers employed by the district eligible for participation was 764 which included 3rd through 8th grade math, reading, and language arts teachers. A representative sample of 46 teachers for review was drawn from the 764 meeting the eligibility requirements.

Data collection

The procedure

The research design was nonexperimental. Information was collected through survey and observation. Participating teachers were asked to complete two surveys, one prior to an observation and a second following the lesson observed. The first survey was collected from all eligible participants. The researcher coordinated with participating teachers chosen for study (the sample) to observe a lesson in its entirety. Following the observed lesson, the teacher completed the second survey, a reflective survey, which

provided additional information on the lesson observed and the teacher's perception of her instructional practices.

The teacher survey

The instrument. The teacher survey was an extensive survey created to capture various elements determined significant to the design of the study. The five sections of the survey included: (a) classroom techniques and experiences; (b) professional development; (c) certifications and endorsements; (d) college preparation and experience; and (e) personal background. These five sections contained items intended to measure perceptions of the relevance and importance of particular aspects of teaching as well as the frequency of use of specific methods and strategies. Each of these concepts was connected to one or both domains of the study. Other sections on the survey, such as personal background and college preparation and experience, were collected and used to describe the participants in the study. Table 1 explains the integration of the five sections of the survey to the study's domains. Note that the fifth section of the survey, personal background, does not fit into either domain but will serve to describe participants involved in the study. The teacher survey is included in the appendix.

Table 1

The survey connected to the two domains of study

Section of survey	Domain of study
(a) Classroom techniques and experiences	Planning and preparation
(a) Classroom techniques and experiences	Instruction
(b) Professional development	Planning and preparation
(c) Certifications and endorsements	Planning and preparation
(d) College preparation and experience	Planning and preparation
(e) Personal background	Descriptive only

The teacher survey was created using items from several surveys found in the literature including those from Deborah Ball and colleagues of the University of Michigan Study (1993) (Kennedy, Ball & McDiarmid, 1993), the National Center for Educational Statistics (NCES) teacher preparation and professional development questionnaire (2000) (Parsad, Lewis & Farris, 2000), Center for the Study of Teaching and Policy (CTP) Teacher and Principal Survey (2002), Bay Area School Reform Collaborative (BASRC) Teacher Survey, Horizon Local Systemic Change (LSC) Math Questionnaire (2002-2003) (Weiss, Pasley, Smith, Vanilower & Heck, 2003), Horizon National Math Survey (Weiss, Pasley, Smith, Vanilower & Heck, 2003), NCES's Schools And Staffing Survey (SASS) teacher questionnaire, and Science Work Experience Programs for Teachers (SWEPT) pre-program survey. Table 2 provides information on these instruments as they relate to item numbers and the domains of this study.

Table 2

Item construction for teacher survey

Instruments referenced	Item Number(s)	Domain of Study
CTP Teacher Survey	2-20	Instruction
BASRC Teacher Survey	2-20	Instruction
Horizon Research, LSC Math, K-8	21-30	Instruction and Planning & preparation
Horizon Research, LSC Math K-8	31-50	Instruction
Horizon Research, Nat'l Math Survey	51-56	Planning & preparation
BASRC Teacher Survey	57-58	Instruction and Planning & preparation
BASRC Teacher Survey	60-67	Planning & preparation
BASRC Teacher Survey	69-77	Planning & preparation
Horizon Research, LSC Math K-8	78-95	Planning & preparation
BASRC Teacher Survey and Michigan Study, Section C	96-105, 181-203	Instruction and Planning & preparation
BASRC Teacher Survey and Michigan Study, Section C	106-119	Planning & preparation
CTP Teacher Survey	168-180	Planning & preparation
CTP Teacher Survey and SWEPT Post-Program Survey	204-263	Planning & preparation
BASRC Teacher Survey	264-270	Planning & preparation
CTP Teacher Survey	279-286	Planning & preparation
NCES SASS Teacher Questionnaire	288-299	Planning & preparation

Distribution procedure. The teacher survey was distributed to all teachers meeting the criteria for participation in the study through the district's internal mailing system. Teachers received a packet that included, along with the teacher survey, a cover letter describing the study and an instruction sheet informing participants of what was being asked of them, of confidentiality issues and compensation information. Also included in this packet were letters of support from the superintendent and the local education association addressing the importance of the study encouraging their participation. Packets were sent out at the same time to all potential participants. A deadline of two weeks was given for individuals to return completed forms to be eligible for

consideration in the study. One hundred seventy-four (23%) completed surveys were returned.

Selecting participants for study from the pool of completed surveys. A list of all participants completing the survey was sent to SAS Institute, an agency managing and warehousing teacher and student data for the state of Tennessee. This agency has access to teacher effect data, along with access to other pertinent demographic information such as subject matter, grade level taught and current school in which the prospective teacher participant teaches. The agency evaluated the information on those completing and returning the survey, considered their level of effectiveness and presented the researcher with a sample, keeping in mind that representation was needed from high (very effective), average and (low) ineffective teachers. Attempts were made to select more participants with high and low effect scores with a smaller number possessing average effect scores. This selection increased the researcher's ability to examine the effective and less effective groups of participants across the continuum of effect scores. This non-random, stratified sample (n=46) was presented blind to the researcher with effect scores of teachers being held until all surveys and observations were completed.

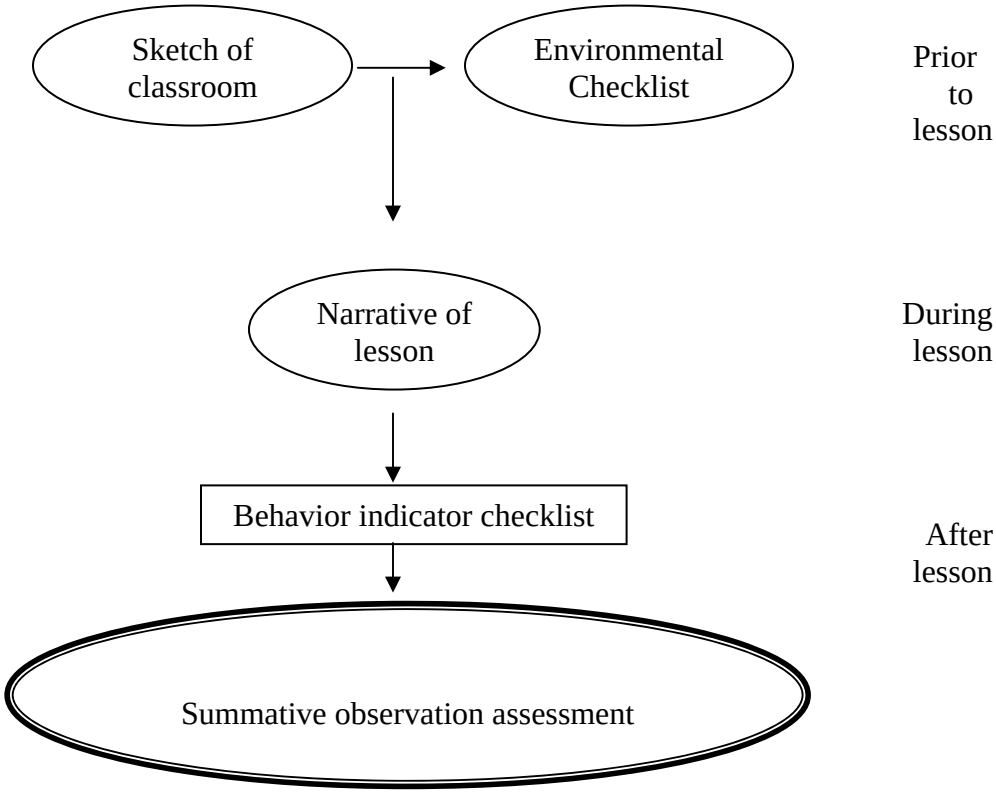
Observational Protocol

A second phase of the study included scheduled observations and reflective surveys. Teachers selected and confirmed for this phase of the study were contacted and scheduled for a classroom observation of one complete lesson. Several instruments were used to capture information during the observation, including a sketch of the classroom observed, a classroom environmental checklist, a narrative of the lesson, a behavior

indicator checklist, followed by a culminating summative observation assessment. The reflective surveys were given to the teacher following the observed lesson. Figure 2 represents the observation process. An example of each instrument is included in the appendix.

Figure 2

The observation protocol



The instruments. The sketch was a hand-drawn representation of the classroom that included information that might be useful to the researcher in capturing particular aspects of the classroom setting. The observer drew the seating arrangement of the students, the location of door(s), windows, teacher's desk, centers and any other artifacts that would best describe the classroom for someone not present in the room at the time of

the observation. The observer was asked to complete the sketch prior to observing the lesson.

The environmental checklist is a form that also was completed by the observer prior to the beginning of the observed lesson. This checklist included specific aspects that might affect the classroom learning environment. The observer was asked to note the presence and condition of the desks, tables, surroundings, books and shelving, and other resources available such as computers, printers, labs, centers, etc.

The narrative contains structured notes taken by the observer during the lesson observed. The observer was trained to capture aspects of (a) what the teacher was doing, (b) what the students were doing, (c) what materials were being used, (d) the academic content of the lesson, and (e) how the students were grouped during the entire lesson.

The behavior indicator checklist is a form that was completed following the lesson observed. This checklist included specific aspects of the lesson that should have been examined/considered by the observer. Each observer was trained to be familiar with these behaviors prior to beginning the observations. If a specific behavior was observed, the observers were instructed to mark it on the checklist. If not observed, the behavior would remain unmarked.

The summative observation assessment is the final form to be completed by the observers. This culminating instrument was meant to combine all observational instruments into a final, and summative, form. It was to be completed soon after the observation, while specifics of the lesson were still fresh on the mind of the observer. This instrument was to be completed *only* after all other observational instruments were completed, using these instruments as evidence of behaviors and practices observed.

Training. The complexity of the observational protocol necessitated extensive observer training. An overview of the process was presented to observers. Following the overview, each instrument was systematically introduced to the observers, progressing through the summative observation assessment.

First, the sketch of the classroom and the environmental checklist was introduced to the observers simultaneously. A video recording of a classroom was shown and observers were instructed to reproduce the classroom in their sketches. The environmental checklist also was completed using the video recorded lesson. The researcher discussed with observers specific aspects of the classroom that should have been drawn and indicated on the environmental checklist. Discussion about the observers' results continued until each observer was comfortable and competent in completing these instruments.

The narrative for each lesson observed also was practiced using videotaped lessons. Observers were asked to watch and script sections of video. As with the sketch and checklist, the researcher discussed specifics of that lesson and observers were asked to comment on how they captured this information. The observers were given the liberty of scripting as they wished as long as they justified and produced written evidence of what was happening during the observed segment. Each segment of the narrative was to contain the five elements previously mentioned: (a) what the teacher was doing; (b) what the students were doing; (c) the content observed; (d) the materials being used; and (e) how students were grouped. Once each observer was comfortable scripting the recorded sections of lessons, the training continued with the behavior indicator checklist.

The indicator checklist is a form that was completed by the observers following the lesson observed. Observers were instructed to complete this form as soon after the lesson observed as possible. The form contains lists of various behaviors that may or may not have been present in the lesson. Observers were asked to note behaviors that were present during the lesson and leave those that were not observed unmarked.

With the sketch, environmental checklist, narrative and indicator checklist completed, the observers were asked to complete the summative observation assessment (SOA). They were directed to find a quiet place where they would not be disturbed [to complete this form]. The researcher emphasized to the observers that the SOA should be completed as soon as possible following the observed lesson and must be completed prior to observing another lesson. The SOA was to be completed using all forms completed prior to, during and following the lesson observed as support for their response.

This exhaustive observation protocol was practiced by observers prior to entering the classrooms for observations. Four videotaped classroom lessons were used to train observers. The number of items in agreement was recorded and a percentage of the total in agreement was calculated. The total number of items was then divided by the number recorded in agreement. The first test agreement was calculated at 81%. The percentage increase to 92.6% for the second test, 84.4% for the third and 93.8% for the final test. Observers practiced using these videotaped recordings until a reliability of 88% was achieved.

Reflective survey

Immediately following the lesson observed, the observers were asked to give the observed teacher a folder containing two reflective surveys to be completed and returned. A self-addressed, stamped envelope was included for return, along with instructions asking the teacher to complete the reflective surveys within 72 hours (or as soon as possible) of the lesson observed. These surveys were designed to capture further information from the teacher pertinent to the subject observed as well as particulars to the lesson observed.

Each teacher observed was asked to complete a general reflective survey**.** This survey was designed to capture additional information specific to the lesson observed, from the teacher's perspective. Teachers were asked to consider the lesson observed when answering the survey questions. Information such as the objective taught, challenges, and strategies used was collected on this form.

Each teacher was asked to complete a survey specific to the subject observed. For example, a teacher observed teaching a math lesson would receive a general reflective survey and a math addendum survey. The subject specific survey was designed to capture specific strategies, philosophies and behaviors as they related to the subject as a whole, not particularly specific to the lesson observed.

Analysis

The data was reduced to a manageable number of variables by employing factor analysis. Analysis was conducted using SAS statistical software. The initial set of factors was obtained using principal component methods and was rotated using the Varimax

method. Varimax, the most commonly used type of factor rotation, permits variables to load more easily into a single factor. The rotated factors were used in the subsequent analyses, regressions and plots.

Items within the five sections of the survey were categorized into the domains of (a) planning and preparation and (b) instruction, as previously mentioned, and were analyzed as such. Factor analysis was employed, presenting a single theme, if appropriate, to represent the related items. These themes were reported as they related to the study's domains. Once factors had been themed, multiple regression analysis was employed to determine relationships between each of the factors and the dependent variable, the teacher effect score.

As previously mentioned, the purpose of most of the instruments was to provide evidence for completing the summative observation assessment form. The sketch, environmental checklist, indicator checklist and the narrative was analyzed for this study. The summative observation form was analyzed in the same manner as the teacher survey, using factor analysis to reduce the variables to workable themes.

The reflective surveys provided both qualitative and quantitative information. The open-ended questions were reviewed, themed and reported as to their relationship with teachers with high and low effect scores. Items with numeric ratings were analyzed in the same manner as the teacher survey and the summative observation instrument.

Items on several of the instruments were collected to provide descriptive information to better define the population of participants included in the study and will be reported as such.

Limitations

A survey was used to collect data on the teachers' perceptions of various aspects of their instruction, beliefs and practices. Teachers are not always aware of their behaviors in the classroom and often articulate that they are doing something that may not be observed in the same way by others. For example, a teacher may think that he gives students ample wait time when asking a question. On a survey, this teacher might rate themselves as doing this well. However, upon viewing a recording of the behavior, she may find she did not give as much time as first thought. The researcher recognizes this limitation and will cautiously report survey findings as perceptions, not proven practices and/or behaviors.

The observational protocol presents another limitation in the amount of time allotted to observing instruction. Given the scope and resources of the researcher, only one observation per participant was possible. The researcher is aware that for a more accurate and thorough view of what is happening in the classroom, more than one observation is needed. Given the exploratory nature of this study, the usefulness of the findings in changing policy are limited. However, the findings will provide constructive information for future observational research of classroom behaviors and their relationship to increased student achievement gains.

The generalizability of the findings presents another limitation. The study is not generalizable beyond the participant pool. The study's purpose was to begin exploring possible relationships between varying levels of teacher effectiveness and the planning, preparation and instructional practices of teachers examined. The findings will be useful for this purpose.

The teacher effect score on student achievement progress is from the year prior to the observations, thus the classroom data is not based on the same classes for which the teacher effect scores were computed. Usually effects are based on causes that occur earlier. This is reversed in this study, although we will proceed as though the reversal is not an issue. This is justified because the high repeatability of the teacher estimates between adjacent years.

Definitions of terms

Academic achievement: Accomplishment and/or success produced through schooling made evident in learning. Academic achievement in most educational arenas is defined by a standardized test given at various junctures in a student's career. It is represented by a score given the student on a standardized test. This score represents the accomplishments of the test taker, the student, on the test at a particular point in time.

Content knowledge: For this study, content knowledge is defined as subject matter knowledge. The term references the teacher and the amount of content knowledge they possess. This is more easily measured in upper grades in that most teachers in these grades are required to have a major or minor in the subject they teach. Elementary teachers that teach all subjects are often generalists and have no specific subject in which they major or minor.

Gain: Gain will be defined as the academic growth or the amount of impact attributed during a specific period of time. To calculate the amount of gain, the equation can vary from a simplified equation of subtracting last year's scale score from this year's scale score to a much more complex and sophisticated model involving much more information than two scale scores. For this study, gain will be defined as the amount of student growth attributed to the teacher in terms of a positive or negative teacher effect score, more specifically, using the Tennessee Value-Added Assessment model. High gains will equate with an effective teacher while low gains will equate with a less effective or ineffective teacher.

Instructional practices: For this study, instructional practices referred to one of two domains of study. The domain was defined by the actions and behaviors carried out in the

classroom. This included teacher actions, classroom techniques and/or events that take place within the classroom during the lesson.

Planning and preparation: Although the two terms are not synonymous, for this study the two represented a single domain. This domain was adopted from Charlotte Danielson's *Enhancing Professional Practice: A Framework of Teacher* (1996) and included the following behaviors: (a) demonstrating knowledge of content and pedagogy; (b) demonstrating knowledge of students; (c) selecting instructional goals; (d) demonstrating knowledge of resources; (e) designing coherent instruction; and (f) assessing student learning. This study adopted the domain as defined and presented in Danielson's framework.

A teacher effect score is used to define the level of effectiveness of a teacher. The score is calculated from the gains of students within the teacher's classroom. This gain is converted to a standard deviation and translates to an effect score. For example: A teacher with an effect score of +2 standard deviations is a more effective teacher at producing student gains than a teacher with an effect score closer to 0 (representing the average score) and a teacher with an effect score of –2 (very ineffective). The teacher effect score is the amount of learning credited to a teacher, in relation to the classrooms' gains on academic achievement tests.

Teacher effectiveness/ineffectiveness: Teacher effectiveness, defined by Flanders & Simon (1969), is "an area of research which is concerned with relationships between the characteristics of teachers, teaching acts, and their effects on the educational outcomes of classroom teaching" (p. 1423). Teacher effectiveness will be measured and/or determined

by the teacher effect score. This effect score is presented as a standard deviation. Teachers above 0 are more effective than those with a teacher effect score lower than 0.

TVAAS: Tennessee Value-Added Assessment System is a "statistical process which provides measures of the influence that school systems, schools and teachers have on indicators of student learning" (Sanders & Horn, 1994). This system determines academic growth over time. The model is the methodology designated by the state of Tennessee to ascertain the effectiveness of its school systems, schools and teachers in producing academic growth among Tennessee students (Sanders & Horn, 1998).

Value-added: The *value added* is that amount of impact that can be accounted for and credited to the various influences on the student's academic achievement gains.

CHAPTER IV

RESULTS AND DISCUSSION

Findings and results from the surveys and observational instruments are reported separately. The demographic section of the following pages will present findings from the categorical data of the teacher survey. No further statistical analyses were run on these data. The intent of these data was to present a more descriptive picture of the participants of the study. Other variables of the teacher survey are discussed further, examining results from the factor analysis. Following the teacher survey, results from the observation instrument are discussed followed by a brief synopsis of findings from the reflective survey.

Demographic data on the participants

Surveys were sent to 764 teachers that met the criteria for participation, which included having taught math, language arts, and/or reading in grades three through nine. One hundred forty-seven returned completed surveys and were included in the pool of possible participants. The sample studied included 46 teachers from that pool of final participants. The 46 were a purposive sample chosen to represent a range of teacher effect scores from critically low to extremely high. Demographic information on this group is presented in the following paragraphs.

Most of the teachers (68%) received degrees from Tennessee institutions, with 47% graduating from the University of Tennessee at Chattanooga located in Hamilton

County, Tennessee. Sixty percent hold a Master's degree. The majority of participants (97%) were female.

Teacher effect scores of this group ranged from -9.18 to 7.14 (see table 3), with the mean and median teacher effect scores of 0.42 and 0.49, respectively. Figure 3 provides the distribution of teacher effect scores among participants in the study. The teaching experience for participants in the study ranged from four to thirty-six years. The categories, as defined in table 4, will be used throughout the analysis to describe groups of teachers by varying levels of experience.

Table 3

Range of teacher effect scores

N	Valid	46
	Missing	0
	Mean	.4279
	Median	.4904
	Std. Deviation	2.779
	Variance	7.724
	Range	16.33
	Minimum	-9.18
	Maximum	7.15

Figure 3

Distribution of teacher effect scores

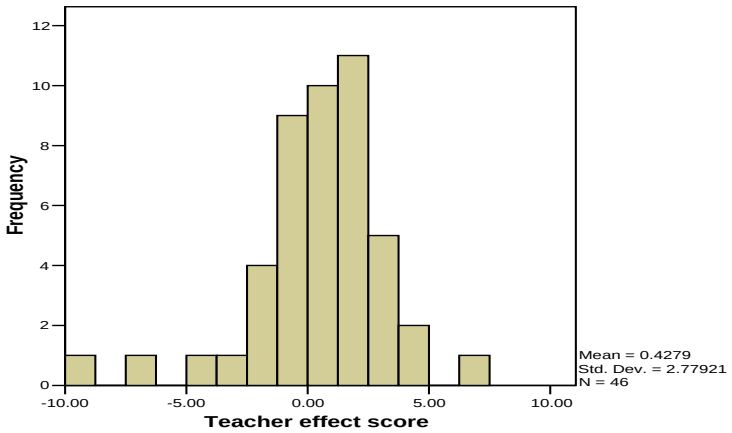


Table 4

Number of years taught

	Years experience	Range of teacher effect score	N	Percent	Cumulative percent
Group 1	less than 5 years	-9.180 to 1.75	12	19.35	19.35
Group 2	5-10 years	-0.930 to 4.04	9	14.52	33.87
Group 3	11-19 years	-0.476 to 3.08	10	16.13	50.00
Group 4	20-25 years	-1.380 to 7.14	14	22.58	72.58
Group 5	26-36 years	-7.317 to 3.19	17	27.42	100.00
Total			62	100.00	

In examining the teacher effect scores, considering years of experience, the lowest teacher effect score is represented among those with less than 5 years of experience. The highest teacher effect score is represented in Group 4 (20 to 25 years experience). The widest range of teacher effect scores is among teachers with the least experience, with the most effective among these having a teacher effect score of 1.75. The least variance in range of teacher effect scores occurs in teachers with 11-19 years experience, Group 3. Ranges of teacher effect scores are presented in table 5.

Teachers in the study represented mainly grades three through six with 8 teachers representing grades 7, 8 and 9. In examining the least experienced teachers (those with the widest range) by range, it can be noted that the highest teacher effect score is found in the 3rd grade while the lowest teacher effect score is among the 8th and 9th grade teachers. Table 5 presents the range of teacher effect scores of each group by grade.

Table 5

Range of teacher effect scores by group for grade observed

Grade	Range of Group 1	n	Range of Group 2	n	Range of Group 3	n
3rd grade	-4.60 – 1.75	6	2.97	1	0	0
4th grade	0.91	1	0.51 – 0.74	2	1.61 – 2.47	3
5th grade	0	0	-0.93 – 4.04	4	-0.47 – 3.08	4
6th grade	-2.32 – 0.81	3	-0.12	1	0	0
7th grade	0	0	0	0	0.43 – 1.81	3
8th & 9th grades	-9.19 - -1.39	2	4.03	1	0	0
Total		12		9		10

Grade	Range of Group 4	n	Range of Group 5	n	Total
3rd grade	-1.38- 3.04	4	-2.64 - -1.19	2	13
4th grade	-0.84 – 0.62	4	-1.19 – 3.19	7	17
5th grade	0.08 – 1.68	2	-2.32 – 1.47	7	17
6th grade	0.44 – 7.14	3	0	0	7
7th grade	0	0	0	0	3
8th & 9th grades	3.15	1	-7.31	1	5
Total		14		17	62

Subjects observed included reading, language arts and math. All but four of the math lessons observed were general math lessons with four teachers observed teaching algebra. Table 6 reports the number of lessons observed in each subject.

Table 6

Subject observed

	N	Percent	Cumulative percent
Reading	17	27.42	27.42
Language arts	17	27.42	54.84
Math	28	45.16	100.00
Total	62*	100.00	

* 46 teachers were observed teaching 62 lessons.

An examination of the subjects observed, taking into consideration the experience groups, finds that the distribution of teachers observed in each subject by group is fairly balanced. The most effective teacher observed teaching reading was from Group 3 (11-19 year experience range) with a teacher effect score of 3.08 and the least effective teacher overall was a Group 1 (less than 5 years experience) math teacher with a teacher effect score of -9.18. The greatest range among the groups occurs in math. More specifically in examining those observed in math, the most experienced teachers (Group 5) had the greatest range of teacher effect scores, followed by teachers with the least experience (Group 1). These and other findings from the review of the table of ranges of teacher effect scores by years experience are presented in table 7.

Table 7

Range of teacher effect scores by group for subject observed

Subject	Range of Group 1	n	Range of Group 2	n	Range of Group 3	n
Reading	-0.47 – 1.75	3	-0.12 – 0.44	2	-0.47 – 3.08	4
Language Arts	-1.83 - -0.91	4	-0.93 – 0.74	2	-0.13 – 1.80	3
Math	-9.18 - -1.39	5	0.46 – 4.04	5	1.61 – 2.47	3
Total		12		9		10

Subject	Range of Group 4	n	Range of Group 5	n	Total
Reading	-1.38 – 1.32	4	-0.72 – 0.17	4	17
Language Arts	-0.84 – 1.68	3	-1.19 – 1.19	5	17
Math	-0.57 – 7.14	7	-7.31 – 3.19	8	28
Total		14		17	62

All participants held Tennessee state certifications, with all but one having a standard teaching certificate. Three were not certified in areas in which they were teaching and one participant did not respond to this statement. A majority (75%) of the participating teachers received their certifications as part of their undergraduate studies. Eight percent earned their certifications through Master's programs, 10% through alternative certification programs after they had begun teaching and 7% earned them through other professional development programs. See table 8 for further detail.

Table 8

State certificate in main teaching assignment field

	N	Percent
part of bachelor's program	47	75.81
part of master's degree program	5	8.06
after I began teaching, alternative program	6	9.68
through continuing professional development	3	4.84
other	1	1.62
Total	62	100.00

When asked to respond to job satisfaction, 71% responded that they were satisfied with their job "most of the time." Another 16% were satisfied with the job "some of the time." Less than 10% were satisfied "all the time" and less than 5% were "almost never" satisfied with their jobs. Table 9 provides results of this finding. Participating teachers' ranges of teacher effect scores by experience categories were examined for the job satisfaction variable. Responses by teachers in Groups 4 and 5 varied more than groups with less experience. Table 10 presents these results and others regarding the range of teacher effect scores by experience.

Table 9

Satisfaction with the job

	N	Percent	Cumulative percent
almost never	3	4.84	4.84
some of the time	10	16.13	20.97
most of the time	44	70.97	91.94
all of the time	5	8.06	100.00
Total	62	100.00	

Table 10

Range of teacher effect score by group for "Satisfaction with the job"

	Range of Group 1	n	Range of Group 2	n	Range of Group 3	n
almost never	0	0	4.03	1	3.08	1
some of the time	0	0	2.97	1	0	0
most of the time	-9.18 - 0.91	11	-0.93 – 4.04	7	0.47 – 2.47	9
all of the time	1.75	1	0	0	0	0
Total		12		9		10

	Range of Group 4	n	Range of Group 5	n	Total
almost never	0	0	1.19	1	3
some of the time	-0.84 – 3.15	4	-2.64 – 0.004	5	10
most of the time	-7.31 – 3.04	8	-7.31 – 3.19	9	44
all of the time	1.68 – 7.14	2	-2.32 – 0.17	2	5
Total		14		17	62

When asked about a first declared major in college, 45% of the participants responded with educational fields. Other responses included (a) art (8%), (b) business and management (8%) and (c) health professions and occupations (6.5%). For other first declared majors given by participants, see table 11.

Table 11

1st declared major

		N	Percent
Valid	All education fields	28	45.16
	Biology/life science	1	1.61
	Chemistry	2	3.23
	English	3	4.84
	Mathematics	1	1.61
	Economics	1	1.61
	Psychology	1	1.61
	Other social science	2	3.23
	Art, fine and applied	5	8.06
	Business and management	5	8.06
	Communications/journalism	1	1.61
	Dance, drama, theater or music	3	4.84
	Health professions and occupations	4	6.45
	All others (general field)	4	6.45
	Total	61	98.37
Missing		1	1.61
Total		62	99.98

Teachers were asked when they decided to become a teacher. Most reported either knowing they “always wanted to be a teacher” (32%) or made this decision once in “college” (30%). Other responses are provided in table 12. Closer examination of these teachers finds that most either with less than 5 years experience (Group 1) or with 20-25 years experience (Group 4) reported that they decided to become teachers in “college.” The majority of teachers in Groups 2 and 3 (range of 5-19 years experience) “always

knew they wanted to be a teacher." Group 5 responses were more varied. Responses categorized by the range of the teacher effect scores by group experience are presented in table 13.

Table 12

When you decided to become a teacher

	N	Percent	Cumulative percent
always wanted to be a teacher	20	32.26	32.36
elementary/primary school	5	8.06	40.32
middle school	3	4.84	45.16
high school	5	8.06	53.22
college	19	30.64	83.86
other	10	16.13	99.99
Total	62	99.99	

Table 13

Range of teacher effect score by group for "When you decided to become a teacher"

	Range of Group 1	n	Range of Group 2	n	Range of Group 3	n
always wanted to be a teacher	-2.32	1	-0.12 – 4.04	5	-0.47 – 2.47	6
elementary/primary school	0.81 – 0.91	2	0.44	1	0	0
middle/junior high school	-4.60 – -1.83	2	0	0	0	0
high school	0	0	0	0	1.73 – 3.08	2
college	-3.99 – 1.75	5	4.03	1	0.43 – 2.30	2
other	-9.18 – -0.78	2	-0.93 – 2.97	2	0	0
Total		12		9		10

	Range of Group 4	n	Range of Group 5	n	Total
always wanted to be a teacher	-1.38 – 0.08	3	-2.64 – 1.47	5	20
elementary/primary school	1.25	1	1.93	1	5
middle/junior high school	0	0	-7.31	1	3
high school	0	0	-2.32 – 0.20	3	5
college	-0.84 – 7.14	8	-6.29 – 1.19	3	19
other	1.32 – 3.04	2	-1.19 – 3.19	4	10
Total		14		17	62

Participants were asked to respond to the various capacities in which they might have served in their schools. More than half (66.13%) responded to serving in a leadership role and/or mentoring new teachers. As many as a third responded to helping develop new curricula (29.03%), supervising new teachers (30.64%), and/or conducting workshops or inservices (38.71%). Responses to these and other capacities in which teachers served are provided in table 14.

Table 14

Various extra capacities served within the school

	N	Percent
served as department chair/grade level chair/team leader	41	66.13
developed or piloted new curricula	18	29.03
formally mentored beginning teacher(s)	32	51.61
supervised student teacher(s)	19	30.64
conducted in-services or workshops for teachers	24	38.71
made observational visits to other schools	13	20.97
made presentations to non-teaching groups (e.g. school board, parents)	21	33.87
conducted individual or collaborative research on a topic of interest to you	27	43.55
represented the school or district on an instructional reform project	8	12.90

When asked about aspects of their preparation for teaching, 76% reported preparation in "how to select and adapt instructional materials." Almost all had preparation (a) in "learning theory or psychology appropriate to the age of the students" they were teaching (92%); (b) in "observations of other classroom teaching" (92%); and/or (c) "feedback on their teaching" (93%).

Sixty percent responded to "10 weeks or more" practice in student teaching. Thirty-four percent reported five to 9 weeks. Four teachers (6.5%) reported "less than four weeks" or "no student teaching." Table 15 reports all responses for student teaching preparation of participants. These responses were disaggregated by experience and range

of experience. All teachers with less than 25 years experience reported some amount of practice in student teaching. Only one teacher, with the least experience and the lowest teacher effect score, responded to having "4 weeks or less" of practice. Responses are presented in table 16.

Table 15

Preparation to include practice (student) teaching

	N	Percent	Cumulative percent
I had no practice	3	4.84	4.84
4 weeks or less	1	1.61	6.45
5-9 weeks	21	33.87	40.32
10 weeks or more	37	59.68	100.00
Total	62	100.00	

Table 16

Range of teacher effect scores by group for "amount of preparation: student teaching"

	Range of Group 1	n	Range of Group 2	n	Range of Group 3	n
I had no practice	0	0	0	0	0	0
4 weeks or less	-9.18	1	0	0	0	0
5-9 weeks	1.75	1	-1.24 – 4.04	3	2.30 – 3.08	2
10 weeks or more	-4.60 – 0.91	10	-9.30 - 4.03	6	-0.46 – 2.47	8
Total		12		9		10

	Range of Group 4	n	Range of Group 5	n	Total
I had no practice	0.82	1	3.19	2	3
4 weeks or less	0	0	0	0	1
5-9 weeks	-1.38 – 3.04	4	-2.32 – 1.93	11	21
10 weeks or more	-0.84 – 7.14	9	-7.31 – 1.47	4	37
Total		14		17	62

Teachers were asked to respond to various statements regarding planning time. Half of the participants (50%) responded as having between 3 and 5 hours per week of

"official" planning time, but only actually being given 1 to 3 hours for planning and preparation (see tables 17 and 18).

Table 17

Official planning time

		N	Percent	Cumulative percent
Valid	1 up to 3 hours	11	17.74	17.74
	3 up to 5 hours	31	50.00	67.74
	5 up to 8 hours	16	25.81	93.55
	8 hours or more	1	1.61	95.16
	Total	59	95.16	
Missing		3	4.84	100.00
Total		62	100.00	

Table 18

Actual time available to you for planning

		N	Percent	Cumulative Percent
Valid	1 up to 3 hours	31	50.00	50.00
	3 up to 5 hours	24	38.71	88.71
	5 up to 8 hours	6	9.68	98.39
	Total	61	98.39	
Missing		1	1.61	100.00
Total		62	100.00	

Seventy-one percent responded 5 to 8 hours as the amount of time they think they need for planning and preparation (see table 19). Most teachers (71%) reported that 5 to 8 hours is the amount of time needed for planning and preparation. Closer analysis examining the amount of time teachers think they need for planning reveals that the responses are similar regardless of level of experience (see table 20).

Table 19

Time needed for planning

		N	Percent	Cumulative percent
Valid	less than 1 hour	1	1.61	1.61
	3 up to 5 hours	8	12.90	14.51
	5 up to 8 hours	44	70.97	85.48
	8 hours or more	9	14.52	100.00
Total		62	100.00	

Table 20

Range of teacher effect by group for "Time needed for planning"

	Range of Group 1	n	Range of Group 2	n	Range of Group 3	n
less than 1 hour	0	0	4.03	1	0	0
3 up to 5 hours	-3.99 – -0.47	3	-0.93	1	0	0
5 up to 8 hours	-9.18 – 1.75	8	-0.12 – 4.04	6	-0.47 – 3.08	7
8 hours or more	-2.32	1	2.97	1	0.43 – 2.47	3
Total		12		9		10

	Range of Group 4	n	Range of Group 5	n	Total
less than 1 hour	0	0	0	0	1
3 up to 5 hours	1.25	1	1.93 – 3.19	3	8
5 up to 8 hours	-1.38 – 3.04	10	-2.64 – 1.47	13	44
8 hours or more	0.44 – 7.14	3	-7.31	1	9
Total		14		17	62

While most teachers report that 5 to 8 hours as necessary to plan and prepare (71%), many reported actually spending more than 8 hours to plan (47%) (see table 21). When examined by experience, similar findings are reported (see table 22).

Table 21

Time spent planning

		N	Percent	Cumulative percent
Valid	1 up to 3 hours	2	3.22	3.22
	3 up to 5 hours	3	4.84	8.06
	5 up to 8 hours	28	45.16	53.22
	8 hours or more	29	46.77	99.99
Total		62	99.99	

Table 22

Range of teacher effect by group for "Time spent planning"

	Range of Group 1	n	Range of Group 2	n	Range of Group 3	n
1 up to 3 hours	0	0	-0.93	1	0	0
3 up to 5 hours	0	0	4.04	1	0	0
5 up to 8 hours	-9.18 – 1.75	5	0.44 - .741	3	-0.47 – 2.30	4
8 hours or more	-3.99 – 0.91	7	-0.12 – 4.03	4	0.434– 3.08	6
Total		12		9		10

	Range of Group 4	n	Range of Group 5	n	Total
1 up to 3 hours	1.68	1	0	0	2
3 up to 5 hours	-0.72	1	1.19	1	3
5 up to 8 hours	-0.84 – 1.25	5	-2.64 – 3.19	11	28
8 hours or more	-1.38 – 7.14	7	-7.31 – 1.17	5	29
Total		14		17	62

Findings from the factor analysis indicators

Categorical and descriptive findings of the teacher survey are reported in the previous section on demographics on participants. The following section will describe findings from the teacher survey specific to the variables that were used in the factor and regression analyses. Variables of the teacher survey were designed to measure specific aspects of the instructional behaviors exhibited by the classroom teacher. Some were reported descriptively (the previous section) but most were analyzed using factor analysis. Following the factor analysis, regression analyses were performed on the factors and the teacher effect scores. Experience levels are considered in this analysis to determine its impact on the relationship between the study's factors and teacher effect scores. These years of experience are treated as categorical variables and grouped as such: (a) Group 1: less than 5 years experience; (b) Group 2: 5-10 years experience; (c)

Group 3: 11-19 years experience; (d) Group 4: 20-25 years experience; and (e) Group 5: more than 25 years experience.

The factor analysis produced 9 major sets of variables or constructs. The variables of these constructs were coded and themed (see table 23). Twenty-seven factors emerged from these nine constructs. Variables included in the construct were chosen if the correlation was 0.60 or higher. Of the 27 factors, not all were found to be statistically significant. Eight main effects across levels of experience and five interactions were found. Twenty-four of the 27 factors were found to be related to years experience defined previously by the five experience groups. These factors will not be reviewed in this section but will be discussed in the final section of this report.

Table 23

Constructs of the teacher survey

Constructs	Variables included
Student engagement in the work	2-20; 31-50
Teacher actions involved in implementation	21-30
Teacher preparedness for specific strategies	51-57; 79-95 (odd numbers)
Importance of various types of assessment	61-67; 69-77; 78-94 (even numbers)
Teacher attitudes towards teaching	96-105; 181-203
Teacher attitudes towards the school environment	106-119
Interaction with peers/colleagues	168-180
Professional development (site, district, personal)	204-236; 240-263
Attitudes of types of professional development	265-269; 279-297

Seven factors were found to have main effects where the factor was a strong and statistically significant predictor of the teacher effect scores across levels of experience. Six additional factors were found to have interactions between the teacher effect scores, the years experience groupings and the factor queried. Each of the interactions and main effects will be discussed in further length in the following pages. Figure 4 provides an

overview of all factors, listing them by construct, their significance and the amount of variance attributed to each factor within that construct. For factors to have statistically significant interactions, the *p* value of the type II analysis had to be less than .05.

Figure 4

Constructs and factors: their significance

Construct	Factor		***Main***	***Interaction****	***% Variance***
1		***Student engagement***			
	1	Students engaged in verbal interactions in the classroom	negative**	2(-);5 (-)	13.58
	2	Students are assessed on tasks lasting more than 1 period	NS	NS	12.63
	3	Students are engaged with materials	NS	1,2(-); 5(+)	10.72
2		***Teacher behaviors***			
	1	Teacher encourages high level interaction	negative	NS	32.69
	2	Teacher follows appropriate lesson sequence	NS	NS	19.75
	3	Teacher assigns homework	NS	NS	12.82
3		***Teacher preparedness for specific strategies***			
	1	Foundations for learning	NS	NS	18.88
	2	Meeting students' needs	NS	NS	16.03
	3	Implementation of performanced-based instruction	NS	4(+); 5(-)	15.00
4		***Importance of various types of assessment***			
	1	Use of varied assessments	NS	5(-)	15.90
	2	Methods to gauge student understanding (learning)	NS	NS	12.76
	3	Evidence of student understanding	NS	NS	12.63
5		***Attitudes towards teaching***			
	1	Teacher attitudes: the environment	NS	1(-); 4(+)	14.88
	2	Teacher attitudes: clear expectations	positive	NS	12.47
	3	Teacher attitudes: ability grouping	NS	NS	7.77
6		***Attitudes towards school environment***			
	1	Teacher attitudes: support for instruction	positive	NS	16.47

	2	Teacher attitudes: planning for teaching content	NS	NS	16.00
	3	Teacher attitudes: grouping for instruction	NS	NS	14.93
7		***Interaction with peers/colleagues***			
	1	Collaborative problem-solving	negative	NS	21.39
	2	Collaboration about teaching and learning	NS	NS	21.27
	3	Peer observation	NS	1(+);4(-)	14.23
8		***Professional development (district, school, personal)***			
	1	School-based professional development	positive**	1,5 (+)	18.60
	2	District-based professional development	negative	NS	12.69
	3	Personal professional development	NS	NS	11.68
9		***Attitudes of types of professional development***			
	1	Professional development characteristics	NS	1(+); 5(-)	19.14
	2	Teachers helping teachers	NS	NS	9.87
	3	Opportunities for collaboration	NS	NS	9.36

**In the interaction column, the number represents the appropriate experience group. The positive or negative symbol in the parentheses represents the direction of the interaction.*

***In cases where there is a significant main effect and a significant interaction, the main effect will be reported.*

NS means not significant.

The interactions. As seen in figure 4, seven factors emerge as having statistically significantly interactions. The five interactions will be discussed in more detail individually. Two of these also were found to have statistically significant main effects and will be discussed in the main effects section following the five interactions.

Factor 3 of Construct 1 is "students are engaged with materials" (see table 25). There is a statistically significant interaction (p= .001) among the teacher effect score and this factor moderated by years of experience (table 25C). This interaction accounts for nearly 50% of the total variance in the teacher effect scores for this construct (see table 25B). Teachers with less than 10 years experience (Groups 1 and 2) have a statistically significant negative slope (p= .03 and p= .01, respectively), while those with 26 or more

years experience (Group 5) have a statistically significant positive slope (p= .002) (see table 25D). The more that teachers in Groups 1 and 2 were reporting such materials being used in their classrooms, the lower their teacher effect scores. The more teachers with the most experience (Group 5) reported the student use of these materials, the higher their teacher effect scores. Results for factor 3 are presented in table 25.

Table 25

Construct 1, Factor 3: Students are engaged with materials

A: ANOVA Summary table for Construct 1, Factor 3, Students engaged with materials

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	134.47	14.94	5.01	< .01**
Error	46	137.12	2.98		
Corrected Total	55	271.60			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.49	377.46	1.726	0.45

C: ANOVA Summary table for Construct 1, Factor 3 and Levels of Experience (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts1_factor 3	1	5.67	5.67	1.90	.08
YrExpGrp	4	105.62	26.40	8.86	<.01**
ts1_factor3*YrExpGrp	4	65.36	16.34	5.48	<.01**

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	-1.06	0.48	-2.17	.03*
ExpGrp2 Slope	-2.66	1.07	-2.48	.01*
ExpGrp3 Slope	0.03	0.55	0.07	.94
ExpGrp4 Slope	-0.66	0.53	-1.26	.21
ExpGrp5 Slope	2.20	0.67	3.27	<.01**
Average Slope	-0.43	0.31	-1.38	.17

(* p <.05, ** p <.01)

The analysis of the third Construct, Factor 3, "implementation of performanced-based instruction," reveals that there is a statistically significant interaction ($p = .001$) between this factor and the teacher effect score, moderated by years of experience (Table 26C). This factor accounts for 50% of the variance in the teacher effect scores for this construct (see table 26B). It is a significant predictor of teacher effect scores only for the more experienced teachers, those in Groups 4 and 5 (see table 26D). Further analysis reveals that teachers in Group 4 (experience levels of 20-25 years) have a statistically significant positive relationship ($p = .01$) with the factor and teachers in Group 5 (more than 25 years experience) have a statistically significant negative relationship ($p = .003$) with the factor. Teachers with 20-25 years experience that report high levels of "implementation of performanced-based instruction," have higher teacher effect scores. The more teachers in Group 5 (more than 25 years experience) reported preparedness in "implementation of performanced-based instruction," the lower their teacher effect scores. For teachers with less than 20 years experience, there is no significant relationship between the "implementation of performanced-based induction" and the teacher effect scores. See table 26 for further analysis.

Table 26

Construct 3, Factor 3: Implementation of performanced-based instruction

A: ANOVA Summary table for Construct 3, Factor 3: Implementation of performanced-based instruction

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	188.35	20.92	5.41	<.01**
Error	48	185.59	3.86		
Corrected Total	57	373.95			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.50	935.47	1.96	0.21

C: ANOVA Summary table for Construct 3, Factor 3: Implementation of performanced-based instruction (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts3_factor 3	1	3.68	3.68	0.95	.33
YrExpGrp	4	111.76	27.94	7.23	<.01**
ts3_factor3*YrExpGrp	4	82.62	20.66	5.34	<.01**

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	-0.92	0.55	-1.68	.09
ExpGrp2 Slope	1.26	0.89	1.41	.16
ExpGrp3 Slope	-0.63	0.88	-0.72	.47
ExpGrp4 Slope	1.09	0.43	2.54	.01*
ExpGrp5 Slope	-2.39	0.77	-3.10	<.01**
Average Slope	-0.31	0.32	-0.98	.33

(* p <.05, ** p <.01)

Closer analysis of Factor 1 in Construct 4, "use of varied assessments," reveals that there is a slight significant interaction (p= .05) (though not significant at the p= <.05 level) between this factor and the teacher effect scores, moderated by years of experience (see table 27C). This factor accounts for 37.5% of the variance in teacher effect scores for this construct (see table 27B). There was a statistically significant negative relationship (p= .03) between teachers in Group 5 (more than 25 years experience) and this factor (see table 27D). The more these veteran teachers reported the use of varied assessments, the lower their teacher effect scores. For other groups of teachers by experience levels, there is no significant relationship between the factor and the teacher effect scores. Table 27 provides the results of the statistical analysis for this factor.

Table 27

Construct 4, Factor 1: Use of varied assessments

A: ANOVA Summary table for Construct 4, Factor 1: Use of varied assessments

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	146.91	16.32	3.34	<.01**
Error	50	244.25	4.88		
Corrected Total	59	391.16			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.37	713.76	2.21	0.30

C: ANOVA Summary table for Construct 4, Factor 1: Use of varied assessments (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts4_factor 1	1	1.55	1.55	0.32	.57
YrExpGrp	4	106.63	26.65	5.46	<.01**
ts4_factor1*YrExpGrp	4	48.84	12.21	2.50	.05

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	-0.80	0.63	-1.27	.21
ExpGrp2 Slope	0.99	0.97	1.01	.31
ExpGrp3 Slope	-0.40	0.86	-0.47	.64
ExpGrp4 Slope	0.76	0.48	1.58	.12
ExpGrp5 Slope	-1.49	0.67	-2.20	.03*
Average Slope	-0.18	0.33	-0.56	.57

(* p <.05, **p <.01)

Closer analysis of Factor 1, “teacher attitudes towards the environment,” in Construct 5 reveals that there is a statistically significant interaction (p= .02) between this factor and the teacher effect scores, moderated by the years of experience (see table 28C). This factor accounts for nearly 42% of the variance of teacher effect scores for this

construct (see table 28B). Teachers in Group 1 (less than 5 years experience) have a statistically significant negative relationship (p= .009) with the factor (see table 28D). The more these new teachers report higher levels of agreement with the statements regarding their educational environment, the lower the teacher effect scores. Table 28 provides further analysis.

Table 28

Construct 5, Factor 1: Teacher attitudes: the environment

A: ANOVA Summary table for Construct 5, Factor 1: Teacher attitudes: the environment.

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	160.31	17.81	3.87	<.01**
Error	48	220.68	4.59		
Corrected Total	57	381.00			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.42	1090.88	2.14	0.19

C: ANOVA Summary table for Construct 5, Factor 1: Teacher attitudes: the environment (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts5_factor 1	1	0.43	0.43	0.09	.75
YrExpGrp	4	86.48	21.62	4.70	<.01**
ts5_factor1*YrExpGrp	4	55.83	13.95	3.04	.02*

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	-3.31	1.22	-2.72	.01*
ExpGrp2 Slope	-0.42	0.46	-0.91	.36
ExpGrp3 Slope	0.32	0.82	0.39	.69
ExpGrp4 Slope	1.31	0.77	1.70	.09
ExpGrp5 Slope	1.41	1.42	0.99	.32
Average Slope	-0.13	0.44	-0.31	.75

(* p <.05, ** p <.01)

Closer analysis reveals that there is a statistically significant ($p = .008$) interaction between Construct 7's Factor 3, "peer observation" and the teacher effect scores, moderated by the years of experience (see table 29C). The factor accounts for 42% of the variance of the teacher effect scores for this construct (see table 29B). For teachers in Group 1, those with less than 5 years experience, those reporting frequently spending time observing other teachers have higher teacher effect scores ($p = .002$). Group 4 (20-25 years experience) is found to have a negative relationship ($p = .03$) with this factor (see table 29D). The more frequently these more veteran teachers report observing other teachers, the lower the teacher effect scores. Table 29 provides the results of the statistical analysis for this factor.

Table 29

Construct 7, Factor 3: Peer observation

A: ANOVA Summary table for Construct 7, Factor 3: Peer observation

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	162.63	18.07	4.11	<.01**
Error	52	228.73	4.39		
Corrected Total	61	391.36			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.41	696.98	2.09	0.30

C: ANOVA Summary table for Construct 7, Factor 3: Peer observation (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts7_factor 3	1	0.75	0.75	0.17	.68
YrExpGrp	4	104.76	26.19	5.95	<.01**
ts7_factor3*YrExpGrp	4	67.12	16.78	3.82	<.01*

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	1.51	0.48	3.16	<.01*
ExpGrp2 Slope	-0.15	0.95	-0.16	.87
ExpGrp3 Slope	-0.17	1.34	-0.13	.89
ExpGrp4 Slope	-1.61	0.74	-2.18	.03*
ExpGrp5 Slope	-0.37	0.44	-0.83	.41
Average Slope	-0.15	0.38	-0.41	.68

(* p <.05, ** p <.01)

There is a statistically significant interaction (p= .01) between Construct 9's factor 1, "attitudes of types of professional development: professional development characteristics" and the teacher effect score, moderated by years of experience (as seen in table 30C). The factor accounts for 41% of the variance of the teacher effect scores for this construct (see table 30B). A positive relationship (p= .01) was found among Group 1 teachers and a negative relationship (p= .01) was found among teachers in Group 5 (see table 31D). Teachers with less than 5 years experience (Group 1) that have higher ratings on their attitudes towards these professional development characteristics also have higher teacher effect scores. For teachers in Group 5 (the most experienced), the higher they rated these statements regarding their attitude towards types of professional development, the lower the teacher effect scores. Table 30 below provides the results of the analysis.

Table 30

Construct 9, Factor 1: Professional development characteristics: Teacher attitudes

A: ANOVA Summary table for Construct 9, Factor 1: Professional development characteristics: Teacher attitudes

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	121.18	13.46	3.09	<.01*
Error	39	169.72	4.35		
Corrected Total	48	290.91			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.41	1480.16	2.08	0.140

C: ANOVA Summary table for Construct 9, Factor 1: Professional development characteristics: Teacher attitudes (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts9_factor 1	1	0.65	0.65	0.15	.69
YrExpGrp	4	53.19	13.29	3.06	.02*
ts9_factor1*YrExpGrp	4	62.08	15.52	3.57	.01*

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	1.98	0.80	2.46	.01*
ExpGrp2 Slope	-0.92	1.14	-0.80	.42
ExpGrp3 Slope	-0.64	1.10	-0.58	.56
ExpGrp4 Slope	0.35	0.56	0.62	.53
ExpGrp5 Slope	-1.53	0.57	-2.68	.01*
Average Slope	-0.15	0.39	-0.39	.69

(* p <.05, ** p <.01)

Summary. Seven factors were found to have interactions between the factor, the teacher effect score and years of experience. Of those seven, two were also found to have main effects and will be discussed in the following section. Most of the interactions found clustered around beginning teachers and those with the most experience. There was very little, if any, interaction occurring between the factors, the teacher effect scores and years of experience for teachers representing Groups 2 and 3 (5-19 years of experience). Of interest in the interactions, most times the more experienced teachers' and beginning teachers' responses were contrary to each other. The exception to this was found in Construct 8, factor 1 where both beginning and more experienced teachers were found to have teacher effect scores positively related to higher agreement on school-based professional development variables.

The main effects. Table 31 presents the findings for Construct 1, Factor 1, "students engaged in verbal interactions in the classroom." This factor accounts for 46% of the variance in the teacher effect scores for this construct (see table 31B). The average slope is slightly negative (- .74) (see table 31D) with the significant *p* value of .01 (see table 31C). The negative relationship indicates that the more frequently teachers report "students engaged in verbal interactions in the classroom," the lower their effect scores.

Table 31

Construct 1, Factor 1: Students engaged in verbal interactions in the classroom

A: ANOVA Summary table for Construct 1, Factor 1: Students engaged in verbal interactions in the classroom

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	126.92	14.10	4.48	<.01**
Error	46	144.68	3.14		
Corrected Total	55	271.60			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.46	387.72	1.77	0.45

C: ANOVA Summary table for Construct 1, Factor 1: Students engaged in verbal interactions in the classroom (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts1_factor 1	1	20.43	20.43	6.50	.01*
YrExpGrp	4	25.85	6.46	2.06	.10
ts1_factor1*YrExpGrp	4	48.44	12.11	3.85	<.01*

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	-1.09	0.81	-1.35	.18
ExpGrp2 Slope	-1.08	0.52	-2.09	.04*
ExpGrp3 Slope	-0.22	0.76	-0.30	.76
ExpGrp4 Slope	0.75	0.46	1.64	.10
ExpGrp5 Slope	-2.06	0.63	-3.25	.01*
Average Slope	-0.74	0.29	-2.55	.01*

(* p <.05, **p <.01)

Closer analysis of Factor 1 of the second construct, "teacher encourages high level interactions," reveals in table 32C that there is a statistically significant (p= .01) negative main effect. The occurrence of "teachers encouraging high level interactions" as part of classroom instruction is found to be a significant predictor of teacher effect scores, accounting for 41% of the variance (see table 32B). This negative relationship (the average slope, -.79, in table 32D) suggests that teachers reporting higher frequencies of this behavior were found to have lower teacher effect scores.

Table 32

Construct 2, Factor 1: Teacher encourages high level interactions

A: ANOVA Summary table for Construct 2, Factor 1: Teacher encourages high level interactions

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	159.69	17.74	3.88	<.01**
Error	50	228.65	4.57		
Corrected Total	59	388.35			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.41	817.04	2.13	0.26

C: ANOVA Summary table for Construct 2, Factor 1: Teacher encourages high level interactions (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts2_factor 1	1	26.98	26.98	5.90	.01*
YrExpGrp	4	112.20	28.05	6.13	<.01**
ts2_factor1*YrExpGrp	4	27.02	6.75	1.48	.22

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	0.29	0.86	0.33	.73
ExpGrp2 Slope	-1.81	0.86	-2.10	.04*
ExpGrp3 Slope	-0.67	0.66	-1.01	.31
ExpGrp4 Slope	-0.14	0.53	-0.28	.78
ExpGrp5 Slope	-1.65	0.69	-2.38	.02*
Average Slope	-0.79	0.32	-2.43	.01*

(* p <.05, ** p <.01)

Table 33C reveals is a slight, though not statistically significant, positive main effect (p= .05) for Construct 5, Factor 2, "teacher attitudes towards clear expectations" of the fifth construct. This factor accounts for 35% of the variance in the teacher effect scores for this construct (see table 33B). The more teachers report higher agreement with these statements on clear expectations, the higher the teacher effect scores. Table 33 provides further detail of the analysis.

Table 33

Construct 5, Factor 2: Teacher attitudes: clear expectations

A: ANOVA Summary table for Construct 5, Factor 2: Teacher attitudes: clear expectations

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	136.72	15.19	2.99	<.01*
Error	48	244.27	5.08		
Corrected Total	57	381.00			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.35	1147.71	2.25	0.19

C: ANOVA Summary table for Construct 5, Factor 2: Teacher attitudes: clear expectations (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts5_factor 2	1	20.16	20.16	3.96	.05
YrExpGrp	4	77.08	19.27	3.79	.01*
ts5_factor2*YrExpGrp	4	11.90	2.97	0.59	.67

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	0.64	0.56	1.13	.26
ExpGrp2 Slope	0.81	0.99	0.82	.41
ExpGrp3 Slope	- 0.67	0.70	-0.32	.74
ExpGrp4 Slope	1.32	0.80	1.64	.10
ExpGrp5 Slope	0.81	0.62	1.30	.20
Average Slope	0.67	0.33	1.99	.05

(* p <.05, ** p <.01)

There is a statistically significant positive main effect (p= < .01) between Construct 6's Factor 1, "teacher attitudes towards the school environment [instruction]: support for instruction," and the teacher effect scores (see table 34C). This factor accounts for 40% of the variance in the teacher effect scores for this construct (see table 34B). It is a significant predictor of teacher effect scores and this significance does not vary by years of experience. Teachers with high levels of agreement with the variables included in the support for instruction factor are found to have higher teacher effect scores (see table 34D). Table 34 provides further detail of the analysis.

Table 34

Construct 6, Factor 1: Teacher attitudes: support for instruction

A: ANOVA Summary table for Construct 6, Factor 1: Teacher attitudes: support for instruction

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	157.29	17.47	3.83	<.01**
Error	51	232.52	4.55		
Corrected Total	60	389.82			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.40	664.99	2.13	0.32

C: ANOVA Summary table for Construct 6, Factor 1: Teacher attitudes: support for instruction (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts6_factor 1	1	42.79	42.79	9.39	<.01*
YrExpGrp	4	94.18	23.54	5.16	<.01**
ts6_factor1*YrExpGrp	4	32.46	8.11	1.78	.14

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	1.85	0.81	2.28	.02*
ExpGrp2 Slope	-0.12	0.61	-0.21	.83
ExpGrp3 Slope	0.80	0.64	1.24	.22
ExpGrp4 Slope	0.27	0.65	0.42	.67
ExpGrp5 Slope	2.32	0.96	2.41	.01*
Average Slope	1.02	0.33	3.06	<.01*

(* p <.05, ** p <.01)

Analysis of Factor 1 of Construct 7, “collaborative problem-solving” reveals that there is a statistically significant negative main effect (p= .04) between this factor and the teacher effect scores, moderated by years of experience (see table 35C). This factor accounts for 36% of the variance in the teacher effect scores for this construct (see table 35B). The factor, “collaborative problem-solving” with peers/colleagues is a significant

predictor of teacher effect scores. The higher the levels of collaborative problem solving in interactions with peers/colleagues reported among these teachers, the lower the teacher effect scores (see table 35D).

Table 35

Construct 7, Factor 1: Collaborative problem solving

A: ANOVA Summary table for Construct 7, Factor 1: Collaborative problem solving

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	142.23	15.80	3.30	<.01*
Error	52	249.12	4.79		
Corrected Total	61	391.36			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.36	727.39	2.18	0.30

C: ANOVA Summary table for Construct 7, Factor 1: Collaborative problem solving (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts7_factor 1	1	20.45	20.45	4.27	.04*
YrExpGrp	4	51.01	12.75	2.66	.04*
ts7_factor1*YrExpGrp	4	23.11	5.77	1.21	.31

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	-2.02	1.35	-1.49	.14
ExpGrp2 Slope	-1.14	0.77	-1.47	.14
ExpGrp3 Slope	0.18	0.87	0.20	.83
ExpGrp4 Slope	0.19	0.85	0.22	.82
ExpGrp5 Slope	-1.53	0.65	-2.33	.02*
Average Slope	-0.86	0.41	-2.07	.04*

(* p <.05, ** p <.01)

Closer analysis reveals a statistically significant (p= <.01) positive main effect between Construct 8, Factor 1, “school-based professional development,” and the teacher

effect scores, moderated by years of experience (see table 36C). This factor accounts for nearly 60% of the variance in the teacher effect scores for this construct (see table 36B). The more teachers report spending time observing other teachers, the higher the teacher effect scores (see table 36D). Table 36 provides further results of this analysis.

Table 36

Construct 8, Factor 1: School-based professional development

A: ANOVA Summary table for Construct 8, Factor 1: School-based professional development

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	218.14	24.23	6.47	<.01**
Error	39	146.08	3.74		
Corrected Total	48	364.22			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.59	1292.26	1.93	0.14

C: ANOVA Summary table for Construct 8, Factor 1: School-based professional development (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts8_factor 1	1	36.50	36.50	9.75	<.01**
YrExpGrp	4	174.10	43.52	11.62	<.01**
ts8_factor1*YrExpGrp	4	55.75	13.93	3.72	.01*

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	2.33	0.65	3.56	<.01*
ExpGrp2 Slope	-0.37	0.78	-0.47	.63
ExpGrp3 Slope	0.20	0.63	0.33	.74
ExpGrp4 Slope	0.11	0.56	0.19	.84
ExpGrp5 Slope	2.56	0.79	3.22	<.01*
Average Slope	0.96	0.31	3.12	<.01*

(* p <.05, ** p <.01)

Analysis of Construct 8's Factor 2, "district-based professional development," (see table 37C) reveals that there is a statistically significant (p= <.01) negative main effect between this factor and the teacher effect scores. This factor accounts for 49% of the variance in the teacher effect scores for this construct (see table 37B). District-based professional development is a strong predictor of teacher effect scores. Contrary to school-based professional development, results in which teachers rate district-based professional development variables higher, a negative relationship with the teacher effect scores is found (see table 37D). The higher the ratings of district-based professional development by these teachers, the lower their teacher effect scores. Table 37 provides the results of the statistical analysis for this factor.

Table 37

Construct 8, Factor 2: District-based professional development

A: ANOVA Summary table for Construct 8, Factor 2: District-based professional development

Source	*df*	Sum of squares	Mean Square	F	*p*
Model	9	178.65	19.85	4.17	<.01**
Error	39	185.56	4.75		
Corrected Total	48	364.22			

B

R-squared	Coeff Variance	Root MSE	Tval_2003 Mean
0.49	1456.48	2.18	0.14

C: ANOVA Summary table for Construct 8, Factor 2: District-based professional development (Type III SS)

Source	*df*	Sum of squares	Mean Square	F	*p*
ts8_factor 2	1	36.14	36.14	7.60	<.01*
YrExpGrp	4	91.33	22.83	4.80	<.01*
ts8_factor2*YrExpGrp	4	9.47	2.36	0.50	.73

D

Parameter	Estimate	Standard Error	t Value	Pr > [t]
ExpGrp1 Slope	-1.24	1.19	-1.05	.30
ExpGrp2 Slope	-0.88	1.13	-0.77	.44
ExpGrp3 Slope	-0.83	0.67	-1.24	.22
ExpGrp4 Slope	-2.10	0.87	-2.39	.02*
ExpGrp5 Slope	-0.65	0.61	-1.08	.28
Average Slope	-1.14	0.41	-2.76	<.01*

(* p <.05, ** p <.01)

Figure 4's "Constructs and factors: their significance" provides evidence of the significance found among the main effects and interactions of the teacher survey. Not to be left out is the significance found among these factors when years of experience are held constant. For the 27 factors analyzed, 25 are found to be statistically significant in their relationship with groups by experience.

Summary. Findings from the teacher survey report that for *all* teachers examined, there is a negative relationship with the teacher effect scores in classrooms where teachers report (a) students engaged in verbal interactions; (b) teachers encouraged high level interactions; (c) [teachers] involved in collaborative problem-solving with peers/colleagues and (d) district provided professional development opportunities. There is a positive relationship among all teachers reporting on the teacher survey in the following areas: (a) [teacher] providing clear expectations; (b) [teacher] support for instruction; and (c) school-based professional development opportunities.

Activities reported by teachers that include student engagement in verbal and/or high level interactions are found among teachers with lower teacher effect scores. In classrooms where the teacher provides clear expectations, receives support for instruction and is provided school-based professional development opportunities, teacher effect

scores are found to be higher. Teachers reporting positively to district-based professional development opportunities are found to have lower teacher effect scores.

Other results are found to be significant among groups of teachers defined by their levels of experience. There are several interactions with the factor and the teacher effect score, dependent on years of experience. Most of the significant findings relate to the least experienced group of teachers.

Negative interactions for Group 1(less than 5 years) include the factors “students are engaged with materials” and “teacher attitudes towards the school environment.” In classrooms where teachers report higher agreement or engagement in these areas, the teacher effect scores are lower. For “students are engaged with materials,” teachers with less than 10 years experiences (including Groups 1 and 2) are found to be negatively related to the teacher effect score, meaning teacher effect scores are lower in classrooms with higher levels of student engagement.

Positive interactions for Group 1 include the factors “peer observations” and “professional development characteristics.” Teachers with less than 5 years experience that reported the importance of and/or higher levels of agreement with these factors are found to have higher teacher effect scores.

A few conflicting results were found among the most experienced and the least experienced teachers. Evidence of student engagement with materials is found be negatively related to the teacher effect score for teachers with the least experience. For teachers with the most experience, the opposite is true. Teachers with the most experience that report high levels of student engagement with materials in their classrooms are found to have higher teacher effect scores. The same contradiction is seen

with attitudes towards “professional development characteristics” seen as important to the teacher. Those with the least experience are positively related to this factor while teachers with the most experience are found to be negatively related with the teacher effect scores. A third significant result for the most experienced teachers is with the factor, “implementation of performanced-based instruction.” The most experienced teachers that report higher frequency of such instruction are found to be less effective, i.e. have lower teacher effect scores.

Findings from the teacher survey confirm the complexities involved in the individual perspectives of teachers as they relate to the measurement of teacher effectiveness. Perspectives of instructional behaviors of effective beginning teachers were often found to be in direct contradiction to those same behaviors for more experienced teachers. A particular instructional behavior might be found to relate positively to student learning for one group while negatively related to the other. This discovery however does not suggest that the findings are not significant and crucial to uncovering aspects of teaching that lead to increased student academic gain. It supports the need for closer examination and better measurements to tease out any differences that might exist among experience levels of teachers. This finding also supports the need for varying the investigation to include observational methods for collecting information on the instructional behaviors.

Findings and discussion on the observation process

The observational system used to record data collected in the classroom includes a sequential set of instruments. First, the observer, prior to observing the lesson, sketched

the classroom, collected information on various aspects of the classroom's physical structure. Also, prior to the observation, an environmental checklist was completed, reporting the condition of various materials in the classroom. A narrative was completed during the lesson, collecting specific points determined prior to the observation. Following the observation, the behavior indicator checklist was completed. Observers used the behavior indicator checklist as a guide to specific comments documented in the narrative. After having completed all of these instruments, the observer then completed the summative observation assessment (SOA) form. This form served to combine the evidence collected on all instruments completed prior to, during and immediately following the observation. A discussion of the findings from the SOA is presented in this section.

Twenty-eight statements of the SOA represent the many variables noted on the sketch, environmental checklist, narrative and behavior indicator checklist. The following paragraphs discuss the percentages of behaviors observed during the observation.

In the planning domain, all behaviors were reported observed in at least 50% of the lessons reviewed. Over 90% of the lessons were found to reflect thorough planning and organization and the resources used supported the achievement of instructional goals. See table 38 for all results.

Table 38

Planning

		Percent observed
b.	The lesson reflected thorough planning and organization.	91.93
d.	The resources used in the lesson supported the achievement of the instructional goals.	90.32
a.	The lesson incorporated tasks and interactions consistent with the content observed.	87.10
c.	The instructional strategies and activities observed in this lesson reflected the teacher's understanding of the students' levels of preparedness, prior knowledge and/or learning styles.	87.10
e.	The strategies and activities observed reflected awareness of diversity and equity.	82.26
f.	The lesson encouraged collaboration among the students.	51.61

In the Implementation domain, teachers were found to demonstrate confidence in teaching the content of the lesson in 87.10% of the classrooms observed. Other areas under this domain also were observed often including appropriate pacing of the lessons to meet the students' developmental needs (79.03%), use of appropriate strategies for the content observed (84%) and appropriate use of wait time (67.74%). There was little evidence of metacognition (defined as students reflecting on work, sharing out with partners and setting future learning goals) with only 12.90% of the lessons documenting this behavior. See table 39 for all results.

Table 39

Implementation

		Percent observed
a.	The instructional strategies were appropriate for the content of the lesson observed. (What was being taught was possible through the instructional strategies observed)	83.87
b.	The teacher demonstrated confidence in teaching the content of the lesson.	87.10
c.	The instruction was paced to meet the developmental needs of the students.	79.03
e.	The questioning strategies emphasized appropriate use of "wait time."	67.74
d.	The questioning strategies emphasized effective use of higher order questioning.	43.55
f.	Evidence of metacognition was present.	12.90

In Content, the third domain of the SOA, two behaviors that were most often observed include "content of the lesson was appropriate for all students in this class" and "content specific vocabulary was used and/or encouraged throughout the lesson;" both were observed in 87.10% of the classrooms visited. "Instruction that included connections to the real-world and other contexts when appropriate" was observed in 61.29% of the lessons. See table 40 for all results.

Table 40

Content

		Percent observed
a.	Content of the lesson was appropriate for all students in this class.	87.10
d.	Content specific vocabulary was used and/or encouraged throughout the lesson.	87.10
c.	Student engagement focused on academic ideas that supported the lesson's objectives.	83.87
b.	The teacher demonstrates an understanding of the concepts taught/presented in the lesson observed through dialogue with her students.	70.97
e.	Instruction included connections to the real-world and other contexts when appropriate/possible.	61.29

Behavior statements of the classroom culture domain were mixed with some being documented as observed during most of the lesson reviewed while others were not observed as often. Evidence of intellectual rigor was observed in 33.87% of the lessons. The remaining two behaviors, "teacher encouraged and valued the active participation of all students" and "student interactions reflected appropriate working relationships" were observed in 95.16% and 87.10% of the lessons, respectively. See table 41 for all results.

Table 41

Classroom culture

		Percent observed
a.	Teacher encouraged and valued the active participation of all students.	95.16
b.	Student interactions reflected appropriate working relationships.	87.10
c.	Evidence of intellectual rigor was present (constructive criticism, the challenging of ideas, etc.).	33.87

Assessment was the fifth domain of the SOA. It was expected that not much assessment would be observed, given that teachers were asked to present a lesson for observation of instructional behaviors. Given this, 64.52% of the lessons observed were noted to have teachers using traditional assessments (e.g., pencil/paper tests, questions and answers, etc). Nearly 60% of the lessons observed were found to have teachers assessing by walking around/monitoring. The least observed behavior (8.06%) was the use of non-traditional assessment (e.g., project, portfolios) by teachers. Results of the assessment domain are presented in table 42.

Table 42

Assessment

		Percent observed
c.	Teacher used traditional assessment. (e.g., paper/pencil test, Q & A)	64.52
e.	Teacher assessed by walking around/monitoring.	59.68
a	Students assessed themselves.	43.55
b.	Peer assessment was evident.	16.13
d.	Teacher used non-traditional assessment. (e.g., project, portfolio)	8.06

Behavioral management, the last domain, was addressed as needed, with 33.887% of the lessons observed found to have the teacher addressing inappropriate behavior that was disruptive to the learning environment. See table 43 for results on the behavioral management domain.

Table 43

Behavior management

		Percent observed
a.	Teacher addressed/managed behavior according to behavioral expectations of the class.	82.26
c.	Behavioral expectations were posted/visible in the classroom.	80.64
b.	Teacher addressed inappropriate behavior that was disruptive to the learning environment.	33.87

Further analysis of the SOA was conducted to determine differences in observations among teachers of varying levels of effectiveness. The following tables in the next section present any main effects, experience and/or interaction significance found among these behaviors. Instructional behaviors with statistically significant main effects or interactions are discussed further.

Summary. For each of the domains, it was determined that behaviors commonly regarded as best practices were occurring in classrooms observed. Most behavioral statements were reported as observed in at least 50% of the classes visited, with most statements having been observed in more than 75% of the classes visited. With the majority of teachers observed exhibiting these behaviors, it becomes extremely difficult to make a case that differences were found among levels of effectiveness. Further analysis will discuss the statistical significance of these statements, but one can surmise from the percentage of observed behaviors across classrooms that theses behaviors are not limited to only the classrooms of the most effective teachers.

Analysis from T-tests run on the Summative Observation Assessment (SOA)

Statements from the SOA were rated as (a) observed, (b) not observed, (c) not applicable or (d) don't know. For analysis, variables were recoded as "observed" and "not observed." Variables were coded as "not observed" if they were marked as (a) "not observed," (b) "not applicable" or (c) "don't know." A t-test was run to determine if differences in the means existed between observed and not observed responses. Regression analysis was run to determine any relationship between the variable, the teacher effect score and years experience. Percentages observed were discussed in the previous section. The following section highlights those instructional behaviors that were found to have statistically significant differences in the means of those observed and those not observed.

There was one instructional behavior in the Content domain with a significant finding to discuss. In 42 of 60 classroom observations the instructional behavior "instruction included connections to the real-world and other contexts when appropriate" was observed (see table 44). The difference in the mean teacher effect scores between those observed and those not observed was statistically significant ($p = .01$) (see table 44B) with the group in which the behavior was observed having the higher mean teacher effect score, 0.86 (see table 44A).

Table 44

Group means for Content 3e: "Instruction included connections to the real world and other contexts when appropriate/possible"

A: Statistics on mean teacher effect scores

Variable	3e: Instruction included connections to the real-world and other contexts when appropriate/possible.	N	Mean
Teacher effect score	Not observed	18	-0.96
Teacher effect score	Observed	42	0.86
Teacher effect score	Diff (not observed-observed)		-1.82

B: T-Test

Variable	Method	Variances	*df*	t Value	Pr [t]
Teacher effect score	Pooled	Equal	58	-2.65	.01*

(* p <.05, ** p <.01)

For the assessment domain, there were no statistically significant main effects among the instructional behaviors. There was a statistically significant interaction with the instructional behavior "peer assessment was evident" and the teacher effect score. Peer assessment was evident in 13 of the 62 lessons observed. There was no statistical difference in the means, but there was a significant interaction between the factor and the mean teacher effect score, moderated by years of experience (see table 45B).

Examining the analyses in differences by group experience, an ANOVA was run. This analysis revealed there to be a statistically significant negative relationship (p <= .01) (see table 45C) with Group 1 teachers that were observed using this instructional behavior (see table 45D). Teachers with less than 5 years experience (Group 1) that were observed using this instructional behavior were found to have lower teacher effect scores. However, teachers with more experience (Group 4: 21-25 years experience) were found to have a positive relationship with this behavior. Teachers in Group 4 observed

employing this behavior were found to have high teacher effect scores (see table 45D).

There were no significant findings among other groups. See table 45 for further details.

Table 45

Group means for Content: “Peer assessment was evident”

A: Statistics on group means

Variable	5b: Peer assessment was evident.	N	Mean
Teacher effect score	Not observed	49	0.32
Teacher effect score	Observed	13	0.21
Teacher effect score	Diff (not observed-observed)		0.11

B: T-Test

Variable	Method	Variances	*df*	t Value	Pr [t]
Teacher effect score	Pooled	Equal	60	0.14	.88

C: Two-way ANOVA summary table

Source	*df*	Sum of squares	Mean Square	F	*p*
5b: Peer assessment was evident.	1	0.03	0.03	0.01	.92
YrExpGrp	4	142.49	35.62	7.60	<.01**
YrExpGrp*Behavior 5b	4	52.81	13.20	2.82	.03*

D: Analysis of least square means of groups observed and not observed

Group	Not observed Teacher effect score LSMEAN	Standard Error	Pr > |t|	Observed Teacher effect score LSMEAN	Standard Error	Pr > |t|
1	-1.14	0.68	.10	-5.50	1.53	<.01**
2	1.14	0.76	.13	2.97	2.16	.17
3	1.20	0.76	.12	2.40	1.53	.12
4	0.68	0.65	.29	3.11	1.25	.01*
5	0.07	0.62	.90	-0.67	0.96	.49

(* p <.05, **p <.01)

For the behavioral management domain (table 46), although the t-test shows there is no significant difference among the group means of those observed and not observed, further analysis by groups reveals a significant F-test in the ANOVA table (see table 46B). Analysis of the ANOVA table (see table 46C), the results show that there is a statistically significant main effect. Classrooms in which behavioral expectations were posted were found to have been taught by teachers with lower teacher effect scores. In the classrooms where behavioral expectations were not noted as being found were taught by teachers with higher teacher effect scores. Further analyses shows Group 1 teacher effect scores to be significantly related to this instructional behavior's being observed (see table 46D and 46E).

Table 46

Group means for Behavior Management: "Behavioral expectations were posted/visible in the classroom"

A: Statistics on group means

Variable	6c: Behavioral expectations were posted/visible in the classroom	N	Mean
Teacher effect score	Not observed	12	-0.02*
Teacher effect score	Observed	50	0.37
Teacher effect score	Diff (not observed-observed)		-0.39

B: T-Test

Variable	Method	Variances	*df*	t Value	Pr [t]
Teacher effect score	Pooled	Equal	60	-0.49	.62

C: Two-way ANOVA summary table

Source	*df*	Sum of squares	Mean Square	F	*p*
6c: Behavioral expectations were posted/visible in the classroom.	1	23.54	23.54	5.30	.02*
YrExpGrp	4	145.96	36.49	8.22	<.01**
YrExpGrp*Behavior 6c	4	58.09	14.52	3.27	.01*

D: Analysis of least square means of groups observed and not observed

Group	Not observed Teacher effect score LSMEAN	Standard Error	Pr > |t|	Observed Teacher effect score LSMEAN	Standard Error	Pr > |t|
1	-9.18	2.10	<.01**	-1.20	0.63	.06
2	0.51	2.10	.80	1.45	0.74	.05
3	1.99	1.21	.10	1.21	0.79	.13
4	1.57	1.05	.14	1.06	0.66	.11
5	-1.28	1.21	.29	0.10	0.56	.85

E: YrExpGrp*SOAC6C Effect Sliced by YrExpGrp for Teacher effect score

YrExpGrp	*df*	Sum of Squares	Mean Square	F	*p*
1	1	58.24	58.24	13.12	<.01**
2	1	0.79	0.79	0.18	.67
3	1	1.29	1.29	0.29	.59
4	1	0.74	0.74	0.17	.68
5	1	4.77	4.77	1.08	.30

(* p <.05, ** p <.01)

Summary. A limited number of instructional behavior statements reviewed were found to produce statistically significant results when analyzed. The interactions presented in the preceding section all include interactions with the behavior, the teacher effect and Group 5, the most experienced teachers. Two interactions were found to be negatively related to the teacher effect score. For teachers in Group 5, those observed using varied assessment and classrooms where there was evidence of implementation of performanced-based instruction had lower teacher effect scores than teachers where these

behaviors were not observed. The third interaction involving Group 5 was a positive relationship with the teacher effect score and classrooms where the students were engaged with materials. Interestingly, for that same behavior (students engaged with materials), responses from teachers in Groups 1 and 2 were found to be negatively related with their teacher effect scores. Teachers from Groups 1 and 2 where it was observed that students were engaged with materials had lower teacher effect scores. The last interaction noted involved teachers in Group 4 and the implementation of performanced-based instruction. Although teachers in Group 5 were found to have a negative relationship between this behavior and the teacher effect scores, teachers in Group 4 were found to have a positive relationship.

The dearth of statistically significant findings among the behavioral statements observed should be carefully noted and consideration given to why so many behaviors did not yield more significant results. The impact of this finding will be discussed further in the conclusion of this report.

Analysis, findings and discussion of the general reflective survey

Following the observation, teachers were asked to complete a general reflective survey and a subject specific addendum. When analyzing data, the number of responses in each cell became very small, specifically for the subject specific addendums. To protect the confidentiality of participants, only select findings for the General reflective survey will be presented.

A qualitative approach was taken when analyzing the general reflective survey. Questions were analyzed in the following way. Responses were examined for the overall

group of participants. For example, 76% responded in agreement with a particular statement. Then further analysis was conducted to determine if there were differences in the responses according to the level of experience of the teacher. If percentages were different from the overall group or particularly interesting combinations were found, they were reported. For example, if 76% of the overall group responded in agreement to a particular statement and only 5% of the most experienced teachers in Group 5 responded in agreement while other groups were much higher, this was noted.

The more important of the analyses was the examination of the responses by levels of effectiveness. The levels of effectiveness were categorized as follows: (1) below average teacher effect scores were those with less than -1.-0 standard deviation; (2) average teacher effect scores were those between -1.0 and 1 standard deviations; and (3) above average teacher effect scores were those with higher than 1.0 standard deviation. Responses to items on the survey were examined considering these levels of effectiveness. An example might read, nearly all teachers with above average teacher effect scores responded in agreement to a particular statement while teachers with below average teacher effect scores disagreed with the same statement.

The former analyses were considered on a continuum, using quantitative analysis to determine the statistical significance of the results. For the following analysis, a cutoff was determined to qualitatively speak to some of the perspectives of teachers at varying levels of effectiveness. No levels of significance were determined. These results are reported by comparing percentages of the various categories of teachers (i.e., by the group overall, by experience or by level of effectiveness).

The General Reflective survey

All teachers observed completed a general reflective survey. Instructed to complete within 48 hours of the observed lesson, teachers were asked to reflect on specific aspects of the lesson and aspects specific to the strategies most often used. Questions on the general reflective survey were categorized under five sections: (a) the lesson observed, (b) course specific, (c) instruction, (d) assessment, and (e) other information. Results from each of the sections are reported in the following paragraphs.

The lesson observed. Teachers were asked to respond to the success of the lesson, rating the success from 0 for "not successful" to 7 for "very successful." There were no responses from teachers rating the success of the lesson as less than 4, with the average of all respondents being 5.7. Table 47A represents percentages of teachers responding to each rating recorded.

Examining the success of the lesson by the level of the teacher effect scores, the findings reveal that 41.67% of teachers that had below average effect scores ratings reported the success of their lessons to be a rating of five (table 47B). Fifty percent of the teachers with above average teacher effect scores and 53.85% those with average teacher effect scores rated their lessons as a six (see table 47B). Examination of the success of the lesson by levels of teacher effect scores by groups of experience found the distribution among groups similar (see table 47C). One might conclude from these findings that ratings from teachers on the success of their lessons would probably not be good indicators of teacher effectiveness.

Table 47

A: Overall ratings for the success of the lesson 1-7 with 7 being successful

Rating	N	Percent	Cumulative Percent
1-3	0	0.00	0.00
4	6	9.68	9.68
5	15	24.19	33.87
6	29	46.77	80.64
7	11	17.74	98.38
Missing	1	1.61	99.99
Total	62	99.99	

B: Frequency and percentage by teacher experience groups

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
Rating	n	%	n	%	n	%	n	%	n	%	N
1-3	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
4	1	8.33	1	11.11	0	0.00	3	21.43	1	5.88	6
5	5	41.67	2	22.22	3	30.00	1	7.14	4	23.53	15
6	5	41.67	6	66.67	5	50.00	6	42.86	7	41.18	29
7	1	8.33	0	0.00	2	20.00	4	28.57	4	23.53	11
Missing	0	0.00	0	0.00	0	0.00	0	0.00	1	5.88	1
Total	12	100.00	9	100.00	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage of teachers' ratings of success of the lesson by below average, average and above average teacher effect scores

	Below		Average		Above		Total
Rating	n	%	n	%	n	%	N
1-3	0	0.00	0	0.00	0	0.00	0
4	0	0.00	2	7.69	4	16.67	6
5	5	41.67	7	26.92	3	12.50	15
6	3	25.00	14	53.85	12	50.00	29
7	3	25.00	3	11.54	5	20.83	11
Missing	1	8.33	0	0.00	0	0.00	1
Total	12	100.00	26	100.00	24	100.00	62

Course specific. When asked about the length of the instructional period per day, most teachers described an instructional period as lasting between 36 and 60 minutes (56.45%). Other time periods included eight reporting 35 minutes or less, nine reporting 65-89 minutes, eight reporting 90-120 minutes and two reporting a varied range for their

instructional periods (see table 48A). Further examination by teacher effect scores reports the findings are similar with most teachers, regardless of teacher effect scores responding to instructional periods lasting between 36 and 60 minutes. It should be noted however, that among teachers with high teacher effect scores, 20.83% reported instructional periods of 90 to 120 minutes, with 70.83% reporting less than 90 minutes (see table 48B). When examined by experience, results are similar to the overall group with the majority in each group responding 36 to 60 minutes as the length of the instructional period per day (see table 48C).

Table 48

Length of instructional period per day

A: Frequency and percentage overall to the length of the instructional period per day by below average, average and above average teacher effect scores

Length of time (minutes)	N	Percentage	Cumulative percentage
35 or less	8	12.90	12.90
36-60	35	56.45	69.35
65-89	9	14.52	83.87
90-120	8	12.90	96.77
other	2	3.23	100.00
Total	62	100.00	

B: Frequency and percentage by teacher experience to the length of the instructional period per day by below average, average and above average teacher effect scores

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
Length of time (minutes)	n	%	n	%	n	%	n	%	n	%	N
35 or less	2	16.77	0	0.00	1	10.00	3	21.43	2	11.76	8
36-60	7	58.33	6	66.67	5	50.00	7	50.00	10	58.82	35
65-89	2	16.77	1	11.11	3	20.00	1	7.14	2	11.76	9
90-120	1	8.33	1	11.11	0	0.00	3	21.43	3	17.65	8
other	0	0.00	1	11.11	1	10.00	0	0.00	0	0.00	2
Total	12	100.00	9	100.00	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage responses to the length of the instructional period per day by below average, average and above average teacher effect scores

	Below		Average		Above		Total
Length of time	n	%	n	%	n	%	N
35 or less	3	25.00	4	15.38	1	4.17	8
36-60	7	58.33	17	65.38	11	45.83	35
65-89	1	8.33	3	11.54	5	20.83	9
90-120	1	8.33	2	7.69	5	20.83	8
other	0	0.00	0	0.00	2	8.33	2
Total	12	100.00	26	100.00	24	100.00	62

Instruction. Teachers were asked to respond to how often the various methods were used in working with the class observed. Most teachers reported that some form of lecture or talking to the class as a whole occurred on an “almost everyday” basis (43.55%). Nearly twenty-one percent reported that lecture or talking to their class as a whole only occurred “1-2 times a week” (see table 49A). When examined by years of experience, at least 30% of the responses from every group were “almost everyday” (see table 49B). When examined by level of effectiveness, around half of the responses from teachers with below average and average teacher effect scores responded “almost everyday” (see table 49C). Only 29.33% of teachers with above average teacher effect scores reported using this strategy “almost everyday.” All responses are provided in table 49.

Table 49

Lecture or talk to whole class

A: Frequency and percentage overall of to lecture to whole class by below average, average and above average teacher effect scores

	N	Percentage	Cumulative percentage
everyday	16	25.80	25.81
almost everyday	27	43.55	69.35
1-2 times per week	13	20.97	90.32
1-2 times per month	3	4.84	95.16
never	3	4.84	100.00
Total	62	100.00	

B: Frequency and percentage of to lecture to whole class by years experience

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
everyday	2	16.67	0	0.00	5	50.00	6	42.86	3	17.65	16
almost everyday	6	50.00	5	55.55	3	30.00	6	42.86	7	41.18	27
1-2 times per week	4	33.33	4	44.44	0	0.00	1	7.14	4	23.53	13
1-2 times per month	0	0.00	0	0.00	1	10.00	0	0.00	2	11.76	3
never	0	0.00	0	0.00	1	10.00	1	7.14	1	5.88	3
Total	12	100.00	9	99.99	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage by teacher experience to lecture to whole class by years experience

	Below		Average		Above		Total
	n	%	n	%	n	%	N
everyday	1	8.33	5	19.23	10	41.67	16
almost everyday	6	50.00	14	53.85	7	29.17	27
1-2 times per week	3	25.00	7	26.92	3	12.50	13
1-2 times per month	1	8.33	0	0.00	2	8.33	3
never	1	8.33	0	0.00	2	8.33	3
Total	12	100.00	26	100.00	24	100.00	62

Similar to results from the previously discussed statement on lecturing or talking to whole class, 95.16% responded to using “teacher-led whole-class discussion” at least 1-2 times per week, with most responding “almost everyday” (see table 50A). Similar results are found when data is examined by group experience and effectiveness.

Regardless of level of experience or effectiveness, findings suggest that most teachers report whole class instruction occurring on an everyday or almost everyday basis (see table 50B and 50C).

Table 50

Teacher led whole-class discussion

A: Frequency and percentage overall to teacher led whole-class discussion by below average, average and above average teacher effect scores

	N	Percentage	Cumulative percentage
everyday	17	27.42	27.42
almost everyday	26	41.93	69.35
1-2 times per week	16	25.81	95.16
1-2 times per month	0	0.00	95.16
never	3	4.84	100
Total	62	100.00	

B: Frequency and percentage by teacher experience to teacher led whole-class discussion by years experience

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
everyday	3	25.00	2	22.22	3	30.00	6	42.86	3	17.65	17
almost everyday	4	33.33	4	44.44	4	40.00	6	42.86	8	47.06	26
1-2 times per week	5	41.67	2	22.22	3	30.00	1	7.14	5	29.41	16
1-2 times per month	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
Never	0	0.00	1	11.11	0	0.00	1	7.14	1	5.88	3
Total	12	100.00	9	99.99	10	100.00	14	100	17	100	62

C: Frequency and percentage responses to teacher led whole-class discussion by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
everyday	3	25.00	6	23.08	8	33.33	17
almost everyday	7	58.33	10	38.46	9	37.50	26
1-2 times per week	2	16.67	10	38.46	4	16.67	16
1-2 times per month	0	0.00	0	0.00	0	0.00	0
never	0	0.00	0	0.00	3	12.50	3
Total	12	100.00	26	100.00	24	100.00	62

Responding to how often student-led whole group discussions or presentations occur in the classroom of the students observed, 33.87% of the teachers reported this method occurring "1-2 times per month." Only 1 teacher reported student-led whole-group discussions as happening "everyday" (see table 51A). The majority of teachers reported this method as occurring "1-2 times a week" (32.26%) or "1-2 times a month" (33.87%). When examined by years of experience, it is interesting to note that for Groups 1-4, the majority of responses fall in the "almost everyday" to "1-2 times per month" categories. If examining both tables (by experience and level of effectiveness), one can conclude that teachers with above average teacher effect scores who reported "almost everyday" come from other group experience levels than Group 5 (see table 51B and 51C). This could be explained in that this instructional strategy is a newer concept and teachers with over 25 years experience may not be comfortable or may not have been trained effectively in using this strategy.

Table 51

Student-led whole-group discussion

A: Frequency and percentage overall to student led whole-class discussion by below average, average and above average teacher effect scores

	N	Percentage	Cumulative percentage
everyday	1	1.61	1.61
almost everyday	13	20.97	22.58
1-2 times per week	20	32.26	54.84
1-2 times per month	21	33.87	88.71
never	7	11.29	100.00
Total	62	100.00	

B: Frequency and percentage by teacher experience to student led whole-class discussion by years experience

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
everyday	1	8.33	0	0.00	0	0.00	0	0.00	0	0.00	1
almost everyday	2	16.67	2	22.22	1	10.00	4	28.57	4	23.53	13
1-2 times per week	6	50.00	4	44.44	3	30.00	4	28.57	3	17.65	20
1-2 times per month	3	25.00	3	33.33	5	50.00	2	14.29	8	47.06	21
never	0	0.00	0	0.00	1	10.00	4	28.57	2	11.76	7
Total	12	100.00	9	100.00	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage responses to student led whole-class discussion by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
everyday	0	0.00	0	0.00	1	4.17	1
almost everyday	3	25.00	2	7.69	8	33.33	13
1-2 times per week	4	33.33	11	42.31	5	20.83	20
1-2 times per month	4	33.33	10	38.46	7	29.17	21
never	1	8.33	3	11.54	3	12.50	7
Total	12	99.99	26	100.00	24	100.00	62

The importance of emphasizing how the subject connects to everyday life was shown in the responses on this statement from teachers completing the general reflective survey. An overwhelming majority of teachers (77%) reported they placed major emphasis on connecting the subject being taught to everyday life (see table 52A). Closer analysis by years of experience revealed that teachers with 5-10 years experience found this issue of little importance with all reporting minor or none, 33% and 67%, respectively (see table 52B). However, when examined by levels of effectiveness, responses mirrored those of the overall group with most teachers (at least 67% in each group) responding that showing the importance of the subject in everyday life was of "moderate" to "major importance.' There seem to be no major difference between groups

defined by levels of experience but teachers with 5-10 years experience differed greatly from beginning teachers and teachers with more than 10 years experience (see table 52C).

Table 52

Showing the importance of the subject in everyday life

A: Frequency and percentage overall to showing the importance of the subject in everyday life

	N	Percentage	Cumulative percentage
none	0	0.00	0.00
minor	2	3.23	3.23
moderate	12	19.35	22.58
major	48	77.42	100.00
Total	62	100.00	

B: Frequency and percentage by teacher experience to showing the importance of the subject in everyday life by years experience

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
none	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
minor	2	16.66	0	0.00	0	0.00	0	0.00	0	0.00	2
moderate	5	41.67	3	33.33	0	0.00	1	7.14	3	17.65	12
major	5	41.67	6	66.67	10	100.00	13	92.86	14	82.35	48
Total	12	100.00	9	100.00	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage responses to showing the importance of the subject of everyday life by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
none	0	0.00	0	0.00	0	0.00	0
minor	1	8.33	1	3.85	0	0.00	2
moderate	3	25.00	3	11.54	6	25.00	12
major	8	66.67	22	84.61	18	75.00	48
Total	12	100.00	26	100.00	24	100.00	62

When teachers were asked to respond to how much emphasis they placed on preparing students for taking standardized tests, most teachers (91.94%) responded placing “moderate” to “major” emphasis on this objective. No one reported that no

emphasis was given to preparing students for standardized testing (see table 53A). When examined by years of experience, teachers with the most experience differed in that nearly 71% reported this as “moderately” emphasized. Other experience groups reported similar to the overall group with about half reporting “moderate” and half reporting “major emphasis.” When analyzed by teacher effect score categories, a large proportion of teachers with below average (66.67%) and average (46.15%) teacher effect scores reported “major” emphasis to preparing students for standardized tests compared to 25% of teachers with above average teacher effect scores reporting this behavior as a “major” emphasis (see table 53C). It is difficult to say why the most experienced teachers report more “moderate” emphasis, but one could speculate that with experience, the emphasis is lessened due to exposure to several years of testing and teachers are comfortable with what they are doing and how students are assessed on these tests.

Table 53

Preparing students for taking standardized tests in the subject

A: Frequency and percentage overall to preparing students for taking standardized tests in the subject

		N	Percentage	Cumulative percentage
	none	0	0.00	0.00
	minor	5	8.06	8.06
	moderate	31	50.00	58.06
	major	26	41.94	100.00
Total		62	100.00	

B: Frequency and percentage of preparing students for taking standardized tests in the subject by years experience

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
Rating	n	%	n	%	n	%	n	%	n	%	N
none	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
minor	1	8.33	1	11.1	1	10.00	2	14.28	0	0.00	5
moderate	5	41.66	5	55.55	5	50.00	4	28.57	12	70.59	31
major	6	50.00	3	33.33	4	40.00	8	57.14	5	29.41	26
Total	12	99.99	9	99.99	10	100.00	14	99.99	17	100.00	62

C: Frequency and percentage responses of preparing students for taking standardized tests in the subject by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
none	0	0.00	0	0.00	0	0.00	0
minor	0	0.00	3	11.54	2	8.33	5
moderate	4	33.33	11	42.31	16	66.67	31
major	8	66.67	12	46.15	6	25.00	26
Total	12	100.00	26	100.00	24	100.00	62

Whole-class grouping as a form of instruction was reportedly used by most teachers responding to this statement from a “few times a week” to “everyday.” Nearly 60% reported using whole-class groupings on a daily basis. No major differences were found when examined by years of experience with the exception of Group 5, all responded using whole group instruction “a few times a week” to “everyday (see table 54B). When examined by level of teacher effect scores, teachers with below average teacher effect scores reported 50% moderate emphasis and 50% major emphasis. Most teachers with average teacher effect scores (61.54%) reported using whole group instruction everyday as did 62.50% of the teachers with above average teacher effect scores (see table 54C).

Table 54

Whole-class grouping

A: Frequency and percentage overall of whole-class grouping

	N	Percentage	Cumulative percentage
rarely or never	1	1.61	1.61
a few times a month	0	0.00	1.61
a few times a week	24	38.71	40.32
everyday	37	59.68	100.00
Total	62	100.00	

B: Frequency and percentage of whole-class grouping by years experience

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
rarely or never	0	0.00	0	0.00	0	0.00	0	0.00	1	5.88	1
a few times a month	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
a few times a week	4	33.33	7	77.78	2	20.00	3	21.4	8	47.06	24
everyday	8	66.67	2	22.22	8	80.00	11	78.6	8	47.06	37
Total	12	100.00	9	100.00	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage responses of whole-class grouping by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
rarely or never	0	0.00	0	0.00	1	4.17	1
a few times a month	0	0.00	0	0.00	0	0.00	0
a few times a week	6	50.00	10	38.46	8	33.33	24
everyday	6	50.00	16	61.54	15	62.50	37
Total	12	100.00	26	99.99	24	100.00	62

A small percentage (6.45%) of the teachers reported using individualized instruction everyday. Most reported using this type of grouping a “few times a week” (35.5%) or a “few times a month” (35.5%). Nearly 21% also reported “rarely or never” using individualized instruction (see table 55A). Further analyses shows that 21% of Group 4 teachers reported using individualized instruction “everyday,” but nearly 43% of the same group reported never using individualized instruction (see table 55B). Closer analysis reveals that teachers categorized with above average teacher effect scores had

the most varied responses with 21% responding “never,” 33% responding a “few times a month,” 25% responding a “few times a week,” 16% responding “everyday” and 1 that did not respond. There were no teachers with below average or average teacher effect scores that reported using individualized instruction “everyday” (see table 55C). From this analysis, individualized instruction also is practiced at an individual preference level among teachers. The method of instruction is effective for some but not all. It can be suggested from these findings that effective teachers, i.e. those with above average teacher effect scores, can be as effective even if rarely or never using the same methods as those them a few times a week or daily. Results are provided in table 55.

Table 55

Individualized instruction

A: Frequency and percentage overall of the use of individualized instruction

	N	Percentage	Cumulative percentage
rarely or never	13	20.97	20.97
a few times a month	22	35.48	56.45
a few times a week	22	35.48	91.93
everyday	4	6.45	98.38
missing	1	1.61	99.99
Total	62	99.99	

B: Frequency and percentage of the use of individualized instruction by years experience

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
rarely or never	2	16.67	0	0.00	3	30.00	6	42.86	2	11.76	13
a few times a month	4	33.33	4	44.44	6	60.00	1	7.14	7	41.18	22
a few times a week	6	50.00	5	55.55	1	10.00	3	21.43	7	41.18	22
everyday	0	0.00	0	0.00	0	0.00	3	21.43	1	5.88	4
missing	0	0.00	0	0.00	0	0.00	1	7.14	0	0.00	1
Total	12	100.00	9	99.99	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage responses of the use of individualized instruction by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
rarely or never	1	8.33	7	26.92	5	20.83	13
a few times a month	6	50.00	8	30.77	8	33.33	22
a few times a week	5	41.67	11	42.31	6	25.00	22
everyday	0	0.00	0	0.00	4	16.67	4
missing	0	0.00	0	0.00	1	4.17	1
Total	12	100.00	26	100.00	24	100.00	62

A majority of teachers (69.5%) were confident in their ability to determine the depth, breadth and pace of coverage of material they were teaching (table 56). One teacher did not respond to this statement and one teacher reported being only “slightly confident” (see table 56A). When reviewed by group experience, findings were similar to the overall group (see table 56B). Analysis of responses by level of teacher effect scores shows that a higher percentage of teachers with average and above average teacher effect scores had a higher percentage responded to being very confident with this method. Nearly forty-two percent of the teachers with below average teacher effect scores reported being “very confident” while 69.23% of the teachers with average and 83.33% of the teachers with above average teacher effect scores reported “very confident” (see table 56C). These differences, though not found with experience were evident as seen in the results provided in table 56 reviewing results by teacher effect scores.

Table 56

Your ability to determine the depth, breadth, and pace of coverage of material you are teaching

A: Frequency and percentage overall in the ability to determine the depth, breadth, and pace of coverage of material you are teaching

	N	Percentage	Cumulative percentage
not at all confident	0	0.00	0.00
slightly confident	1	1.61	1.61
moderately confident	17	27.42	29.03
very confident	43	69.35	98.38
missing	1	1.61	99.99
Total	62	99.99	

B: Frequency and percentage by teacher experience groups in the ability to determine the depth, breadth, and pace of coverage of material you are teaching

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
not at all confident	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
slightly confident	0	0.00	0	0.00	0	0.00	1	7.14	0	0.00	1
moderately confident	6	50.00	3	33.33	2	20.00	1	7.14	5	29.41	17
very confident	6	50.00	6	66.67	7	70.00	12	85.71	12	70.59	43
missing	0	0.00	0	0.00	1	10.00	0	0.00	0	0.00	1
Total	12	100.00	9	100.00	10	100.00	14	99.99	17	100.00	62

C: Frequency and percentage responses in the ability to determine the depth, breadth, and pace of coverage of material you are teaching by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
not at all confident	0	0.00	0	0.00	0	0.00	0
slightly confident	0	0.00	1	3.85	0	0.00	1
moderately confident	7	58.33	7	26.92	3	12.50	17
very confident	5	41.67	18	69.23	20	83.33	43
missing	0	0.00	0	0.00	1	4.17	1
Total	12	100.00	26	100.00	24	100.00	62

Summary. In a review of the Course specific and Instruction sections of the General Reflective Survey, it should be noted that most findings were inconclusive in adequately distinguishing differences between teacher effect scores. In some cases there

were small differences, but analysis revealed little evidence to be conclusive in significant results.

Assessment. Teachers were asked to respond to various statements regarding assessment. Many teachers (45.16%) reported using pre-tests before beginning a new unit as a method of assessment to a "slight extent." Eight percent reported never using pretesting and twice that many (16.13%) reported using pretesting to a "great extent" (table 57A). Responses were similar when reviewed by years of experience (see table 57B). A review by teacher effect scores provides some interesting findings. Thirty-three percent of teachers with below average teacher effect scores report to a "great extent" pre-testing prior to beginning a new unit (see table 57C). Compare that percentage to those of the average and above average teachers and one quickly finds that more effective teachers (50% each) report to using pre-tests only to a slight extent. To put this another way, more effective teachers are likely to less often pre-test their students prior to beginning a new unit. All responses are provided in table 57.

Table 57

Pre-tests before beginning a new unit

A: Frequency and percentage overall in the ability to pre-test before beginning a new

	N	Percentage	Cumulative percentage
not at all	5	8.06	8.06
slight extent	28	45.16	53.22
moderate extent	18	29.03	82.25
great extent	10	16.13	98.38
no response	1	1.61	99.99
Total	62	99.99	

B: Frequency and percentage by teacher experience groups in the ability to pre-test before beginning a new

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
not at all	1	8.33	0	0.00	0	0.00	0	0.00	4	23.53	5
slight extent	3	25.00	5	55.55	8	80.00	5	35.71	7	41.18	28
moderate extent	5	41.67	2	22.22	1	10.00	6	42.86	4	23.53	18
great extent	3	25.00	2	22.22	1	10.00	3	21.43	1	5.88	10
no response	0	0.00	0	0.00	0	0.00	0	0.00	1	5.88	1
Total	12	100.00	9	99.99	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage responses in the ability to pre-test before beginning a new unit by below average, average and above average teacher effect scores

	Below		Average		Above		Total
	n	%	n	%	n	%	N
not at all	1	8.33	2	7.69	2	8.33	5
slight extent	3	25.00	13	50.00	12	50.00	28
moderate extent	3	25.00	7	26.92	8	33.33	18
great extent	4	33.33	4	15.38	2	8.33	10
no response	1	8.33	0	0.00	0	0.00	1
Total	12	99.99	26	99.99	24	99.99	62

Short-answer tests were used to a “moderate extent” by the majority of teachers (45.16%) (see table 58A). The findings are similar upon review by years of experience with the majority of teachers in all groups reporting a moderate extent of the use of short-answer tests in their classrooms (see table 58B). When examined by teacher effect scores, a finding similar to that of pre-testing emerges. Teachers with below average teacher effect scores have responded to using this form of assessment more often than teachers with average and above average teacher effect scores (see table 58C). As with pretesting, it appears that teachers with below average teacher effect scores are using this form of assessment more often than are more effective teachers.

Table 58

Short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank)

A: Frequency and percentages overall to the use of short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank)

	N	Percentage	Cumulative percentage
not at all	4	6.45	6.45
slight extent	13	20.97	27.42
moderate extent	28	45.16	72.58
great extent	16	25.81	98.39
no response	1	1.61	100.00
Total	62	100.00	

B: Frequency and percentage in the use of short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank) by teacher experience groups

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
not at all	1	8.33	0	0.00	1	10.00	0	0.00	2	11.76	4
slight extent	2	16.67	2	22.22	4	40.00	4	28.57	1	5.88	13
moderate extent	6	50.00	5	55.55	2	20.00	7	50.00	8	47.06	28
great extent	3	25.00	2	22.22	3	30.00	3	21.43	5	29.41	16
no response	0	0.00	0	0.00	0	0.00	0	0.00	1	5.88	1
Total	12	100.00	9	99.99	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage to the use of short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank) by teacher effect score

	Below		Average		Above		Total
	n	%	n	%	n	%	N
not at all	0	0.00	1	3.85	3	12.50	4
slight extent	1	8.33	5	19.23	7	29.17	13
moderate extent	6	50.00	13	50.00	9	37.50	28
great extent	4	33.33	7	26.92	5	20.83	16
no response	1	8.33	0	0.00	0	0.00	1
Total	12	99.99	26	100.00	25	100.00	62

Summary. Contrary to standards of best practices currently observed, assessment looks much different for effective teachers reviewed in this study. We see from the analysis reported that effective teachers are not necessarily pre-testing students prior to

beginning new units. However, these same effective teachers are not necessarily testing students using conventional methods such as short-answer tests either. Those found to be using short-answer tests to a moderate extent were found to more often have lower teacher effect scores.

On mentoring. About one-third of the teachers (29%) reported being officially mentored as a new teacher, while almost 39% reported being unofficially mentored as a new teacher. Much of the official and unofficial mentoring happened in the grade level (tables 60A and 61). When examining Group 5 responses, it was interesting to note that no teachers in that group reported having been officially mentored (see table 60B). Of those unofficially mentored, the mentoring occurred as a new teacher, at grade level and/or in their school (see table 61B). Review by levels of effectiveness provided little support for mentoring, either official or unofficial. The results of both methods of mentoring, examined by group experience and teacher effect scores, are shared in tables 60 and 61.

Table 60

Officially mentored

A: Frequency and percentage overall of those officially mentored

	N	Percentage
as a new teacher	18	29.03
in your subject area	7	11.29
at your grade level	14	22.58
on your hall	10	16.13
in your school	13	20.97

B: Frequency and percentage of those officially mentored by teacher experience groups

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
as a new teacher	11	91.67	3	33.33	3	30.00	1	7.14	0	0.00	18
in your subject area	4	33.33	1	11.11	1	10.00	1	7.14	0	0.00	7
At your grade level	8	66.67	2	22.22	3	30.00	1	7.14	0	0.00	14
on your hall	7	58.33	1	11.11	1	10.00	1	7.14	0	0.00	10
in your school	8	66.67	3	33.33	1	10.00	1	7.14	0	0.00	13

C: Frequency and percentage of those officially mentored by teacher effect score

	Below		Average		Above		Total
	n	%	n	%	n	%	N
as a new teacher	6	50.00	7	26.92	5	20.83	18
in your subject area	2	16.67	3	11.54	2	8.33	7
at your grade level	5	41.67	4	15.38	5	20.83	14
on your hall	4	33.33	4	15.38	2	8.33	10
in your school	4	33.33	6	23.08	3	12.50	13

Table 61

Unofficially mentored

A: Frequency and percentage overall of those unofficially mentored

	N	Percentage
as a new teacher	25	40.32
in your subject area	4	6.45
at your grade level	12	19.35
on your hall	7	11.29
in your school	14	22.58

B: Frequency and percentage of those unofficially mentored by teacher experience groups

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
as a new teacher	8	66.67	6	66.67	4	40.00	3	21.43	4	23.53	25
in your subject area	2	16.67	1	11.11	1	10.00	0	0.00	0	0.00	4
at your grade level	3	25.00	3	33.33	0	0.00	2	14.28	3	17.65	12
on your hall	2	16.67	2	22.22	1	10.00	1	7.14	0	0.00	7
in your school	5	41.67	4	44.44	2	20.00	2	14.28	1	5.88	14

C: Frequency and percentage of those unofficially mentored by teacher effect score

	Below		Average		Above		Total
	n	%	n	%	n	%	N
as a new teacher	4	33.33	11	42.31	10	41.67	25
in your subject area	0	0.00	2	7.69	2	8.33	4
at your grade level	2	16.67	4	15.38	6	25.00	12
on your hall	1	8.33	2	7.69	4	16.67	7
in your school	1	8.33	7	26.92	6	25.00	14

Forty percent of the teachers reported covering 76-95% of their textbook during the course. Seven reported using no textbook with a cumulative percentage of 27% using less than 50% of the textbook (see table 62A). Examining the results by level of experience, there is a difference between beginning teachers and other groups. Nearly 42% of beginning teachers reported using no textbook (see table 62B). Of those reporting using no textbook, six of the seven had average or above average teacher effect scores. However, all of the 8 responding to using 96-100% of the textbook, also were average or above average (see table 62C). Five beginning teachers reported using no textbook. Interestingly though is that only 1 response was reported as using no textbook in the below average teacher effect table. Thus, of the beginning teachers, the four reporting using no textbook did so with average or above average effect scores. Findings on the use of the textbook are similar to those of individualized instruction. Some teachers are doing it and getting effective results and some are not.

Table 62

Percentage of textbook covered during the course

A: Frequency and percentage overall of percentage of textbook covered

	N	Percentage	Cumulative percentage
no textbook used	7	11.29	11.29
less than 25%	1	1.61	12.90
25-50%	9	14.52	27.42
51-75%	12	19.35	46.77
76-95%	25	40.32	87.10
96-100%	8	12.90	100.00
Total	62	100.00	

B: Frequency and percentage of percentage of textbook covered by teacher experience groups

	Group 1		Group 2		Group 3		Group 4		Group 5		Total
	n	%	n	%	n	%	n	%	n	%	N
No textbook used	5	41.67	1	11.11	0	0.00	1	7.14	0	0.00	7
less than 25%	0	0.00	1	11.11	0	0.00	0	0.00	0	0.00	1
25-50%	4	33.33	0	0.00	4	40.00	1	7.14	0	0.00	9
51-75%	1	8.33	3	33.33	2	20.00	2	14.29	4	23.53	12
76-95%	2	16.67	3	33.33	4	40.00	6	42.86	10	58.82	25
96-100%	0	0.00	1	11.11	0	0.00	4	28.57	3	17.65	8
Total	12	100.00	9	100.00	10	100.00	14	100.00	17	100.00	62

C: Frequency and percentage of percentage of textbook covered by teacher effect score

	Below		Average		Above		Total
	n	%	n	%	n	%	N
no textbook used	1	8.33	4	15.38	2	8.33	7
less than 25%	0	0.00	1	3.85	0	0.00	1
25-50%	3	25.00	3	11.54	3	12.50	9
51-75%	2	16.67	3	11.54	7	29.17	12
76-95%	6	50.00	10	38.46	9	37.50	25
96-100%	0	0.00	5	19.23	3	12.50	8
Total	12	100.00	26	100.00	24	100.00	62

To summarize the general reflective survey. The general reflective survey was administered to capture the viewpoint of the teachers that participated in the study. As noted in many of the responses reported, findings were sometimes connected to specific

groups of teachers (i.e., those within specific experience ranges or teacher effect levels) but overall, most findings could not be separated into sections so easily. The survey's purpose was to support findings from the quantitative analyses run on the teacher survey and the t-tests from the summative observation assessment (SOA). Though obvious and significant findings were not readily apparent in the qualitative results of the general reflective survey, inferences can be drawn to make connections between these responses and the significant findings of the teacher survey and SOA. These connections and conclusions will be discussed further in the final section of this report.

CHAPTER V

IMPLICATIONS, RECOMMENDATIONS AND CONCLUDING THOUGHTS

In the previous chapter, findings were reported under the domains of the study, planning and preparation and implementation of instructional practices. In the examination of the relationship between measures of teachers' instructional behaviors and measures of their students' academic progress, these findings can be summed up in three major points. First, in reviewing a teacher's instructional behaviors and the relationship of those behaviors to student academic progress, some instructional behaviors were found to be significant predictors of the teacher effect scores among all teachers examined in this study. Second, in reviewing these instructional behaviors and the relationship of those practices to student academic progress, some instructional behaviors were found to be significant predictors of teacher effect scores, but only when experience was held constant. Lastly, most instructional behaviors examined in this study were found to be limited and lacking in their ability to predict the teacher effect scores of participants reviewed. The implications of these findings, the recommendations for consideration and concluding thoughts are presented in this final chapter.

Implications

Just as important as the findings are the implications of these findings to what we determine to be "best practices" noted in classroom instruction. Implications of these findings include: (1) maybe there is not AN effective instructional behavior; (2) possibly a sequence of actions is more representative of effective behavior; (3) perhaps clusters of behaviors, instead of individual behaviors, make the difference; (4) perhaps effective behaviors are immeasurable; and/or (5) the approach in the strategy used to capture these findings was inadequate.

It is not the intent of this research to suggest that there is a single behavior that *must* be present among effective teachers. There may not be a behavior that is common among all but maybe there *is* an effective practice. To further explain, it may be that behaviors are nested under some larger umbrella of macro-behaviors and/or actions. The overarching behavior, combined with other specific actions may in turn create an effective practice. The overarching behavior may actually be based in a personality trait. Examining individual behaviors or traits in isolation may not be an appropriate way to determine these differences.

Individual effective behaviors are often elusive and sometimes missed in an observation. Consideration might be given to the practices viewed in a sequence instead of independently. Do effective practices follow some unknown sequential order? Is there an order in which these practices should be learned and/or exhibited in the classroom? Would it be appropriate to teach preservice teachers how to differentiate for various reading levels prior to teaching them how to use various reading strategies independently? One would presume that preservice teachers would need to know

something about the various teaching strategies *before* they would be able to execute them well, especially when differentiating the strategies according to the needs of the students they serve. Would it be appropriate for that teacher to implement these various differentiated reading strategies to students prior to assessing the class to see which strategy would be best for each? Again, the order in which these behaviors are not only taught but the order in which they are implemented in the classroom may contribute to the effectiveness. Maybe the sequence spoken to here is being overlooked. Maybe there *is* an effective practice, but *not* an effective behavior.

The idea of behaviors presented in a particular sequence is interesting. There could be a set of behaviors that when implemented simultaneously may be very effective, but when implemented individually are not. Could it be that if a few of the behaviors in a set are removed and others added, the set of behaviors once deemed effective no longer works? Maybe there is the "cluster" of behaviors that create an effective practice relative to the type of lesson being taught. If this notion were to be considered, how then do we determine what this cluster of behaviors looks like? It could be that over time clusters of behaviors might be captured by examining numerous lessons. This may lead to other questions. Would these clusters then distinguish themselves to particular types of lessons? Would the cluster of behaviors for an introduction to a new idea look the same as would that of a review lesson? Do these questions imply that behaviors, if examined within the context of the lesson and activity being observed, naturally cluster in groups?

It has often been stated that good teachers are born, not taught or trained. Another implication of this study's findings might be that behaviors that are most important are unobservable and perhaps not measurable. Could it be that the difference is in the innate

abilities of the effective teacher to know when, how and where to implement effective behaviors? Could ineffective teachers be going through the same motions mechanically but not linking them in productive ways? This might explain why it was difficult in this study to discriminate distinguishable differences. The often unobservable aspects of teaching, the innate abilities of the individual teacher, may play a major role in determining practices that are most successful at promoting academic growth. Research procedures to date are unable to tap these effectively.

In an effort to systematically capture predetermined behaviors in a scientific and formulated manner, the validity of the instrument could have been jeopardized. Maybe there were effective practices occurring in the classrooms examined, but the instrument used to capture those practices was insensitive to them. Since student learning is the underlying definition of the effectiveness of the teacher in this study, it could be that the student should be the focus of the research. Another alternative might be that we are examining the wrong behaviors altogether. In this study many behaviors that had previously been suggested as best practices that were found to be limited or lacking in their ability to predict differences in teacher effect scores. This may speak to the need to reexamine all aspects of the design to determine if a different plan and system might lead to different results.

Summary of implications. Findings from the study suggest there are several areas that could be further examined and show promise as being predictors of teacher effect scores. It may be that a particular instructional behavior may not fit all teachers, but may be better suited for teachers within a specific level of experience or perspective. Also, it might be that the important differences in these teachers may not be related to

their behaviors, but to something else. Implications from the teacher survey and observation instrument lead to several recommendations.

Recommendations

Given the divergent implications presented in the previous section, recommendations for research can be summed up in three categories. First, consider other factors that might be congruent with best practices. Second, consider the behaviors within the context of the lesson and class involved. Third, investigate the impact of experience on the amount of student learning.

A study examining best practices of the teacher, in isolation of other aspects of the classroom, might be limited in its ability to accurately provide findings that predict positive student learning outcomes. We might find that best practices are a place to start, but from the findings of this study best practices were seen in classrooms of both effective and ineffective teachers. Future research should place more emphasis on the person(s) being examined. A design that incorporates study of the teacher as an individual, his/her personality as it relates to how (s)he teaches may be helpful in teasing out differences between those that are effective and those that are ineffective.

It also might be beneficial to consider students in the design frame. If the outcome is the academic growth of the student, aspects of how the student learns, interactions that are found when the learning occurs and relationships that foster such learning should not be minimized. There is a dearth of previous research on teacher-student interactions. This research should be re-examined for its efficacy and possible use in a modern study of interactions, using the teacher effect score as the predictor variable.

A more contextualized observation system would consider the teacher and the student in various aspects of the classroom: teaching; learning, relationships and context. It is possible that personality traits and other characteristics which teachers bring to the classroom are being overlooked, or at least minimized, and should be carefully studied before discounted.

Second, simply capturing that a behavior occurs, apart from the context of that lesson, is not sufficient. The appropriateness of the behavior exhibited also should be more closely examined. Findings of this study suggest that some behaviors were independently strong predictors of teacher effect scores while others were dependent on experience and others showed no relationship at all to teacher effect scores. This implication requires the researcher to step back and examine the process of data collection.

As suggested by Barr (1929), the appropriateness of the context seems to play an important role when considering factors that relate to teacher effectiveness. For example, in Barr's study, teachers were asked to present a lesson. An observer documented practices and behaviors exhibited in this classroom. However, the lesson's content was not prescribed or even suggested for that matter. Some teachers responded that the focus of the lesson was review, others as an introductory lesson, while others reported lessons as clarifications or further discussions on previously discussed topics. These differences in delivery could be compared to Barr's consideration of "appropriateness within the context" (Barr, 1929). Teachers would likely not employ the same practices and strategies in an introductory lesson that they might once the material has been introduced and students are given an opportunity to practice what has been taught. A limitation of

this study (no defined lesson) would strongly suggest that lessons be observed within a specific context (i.e., introduction, practicing concepts, reviewing ideas, etc.). A more carefully designed study that considers the appropriateness of the behaviors within the context of the lesson observed might better capture distinguishable differences in instructional practices.

Third, for 25 of the 27 factors of the teacher survey there was a statistically significant difference among the five groups of experience analyzed. The two factors in which experience was not significant include “students engaged in verbal interactions in the classroom” which had a statistically significant negative main effect and attitudes of “teachers helping teachers” which was not found to have any significance. Throughout the findings of this study, experience played a significant role. To consider the instructional behaviors absent the experience of the teacher is clearly misleading. Any further study of teacher effectiveness should consider this seriously.

Incorporating suggestions from recommendations one through three, researchers should consider creating a new instrument that would be better at capturing differences in teachers that are most effective and those that are ineffective in terms of student academic gains measured by valued-added scores. If these behaviors are indeed measurable, the need for a more scientifically designed observational instrument is crucial to accurately capture variables that might be predictors of teacher effectiveness. The current process has proven ineffective. A worst case scenario would be one in which researchers become content with the findings of date and push for policies to change the way we prepare teachers based on a flawed design. This study suggests that there is little basis on which to push for any specific behaviors.

Much of the literature on teacher effectiveness comes from the extensive body of literature on teacher evaluation (Beecher, 1949; Charters & Waples, 1929; Danielson, 1996; McBer, 2000; Merriam, 1905). These evaluations may be the problem in the design of many studies on teacher effectiveness. Most teacher evaluations are the result of observation. This observational information then is crafted into formal rating systems that are used in research studies to examine teacher effectiveness. These rating and ranking systems are met with apprehension when used as instruments to define teacher effectiveness. As evaluations, they are met with criticisms of subjectivity and bias. As observation instruments, understandably, they are confounded by what was previously known as best practices. With the academic gains (i.e., student learning) as the outcome measure for the predictor, effective teaching, it is imperative that a valid instrument be used to capture student learning. Careful attention should be given to what is happening in the classroom. It is not sufficient to use a checklist to capture a group of best teacher behaviors that could or could not be predictors of academic success for the students that are recipients of what we have determined to be "best practices."

Particularly in the age of accountability, with pay for performance knocking on the doors of many states, the use of a proven, robust observational instrument, or combination of companion instruments is imperative not only to validate the assessments adopted by the policymakers, but to provide fairness and equity to the educators being evaluated by this instrument. It is time to reconsider our observational assessment processes and work to craft an instrument that is more successful in capturing interactions and classroom delivery of instruction which lead to greater student learning.

Concluding thoughts

Continued efforts are needed to address the deficits of observational instruments used in determining the effective practices and behaviors in the classroom. It is apparent that findings of this study are held within the confines of the observation variables reviewed. Is it possible that other variables might be better at predicting teacher effectiveness? Have these variables been determined, researched and/or considered in such research? It would be careless to report that there are no predictors of teacher effectiveness except those found in this study.

It is the belief of the researcher that an adequate observational instrument has not been found. Much work has been done and we have come a long way. Yet more work is necessary before we put children in jeopardy of being recipients of poorly constructed instructional strategies that are not founded on a series of sound and rigorous scientific inquiries. Without a clear connection between teaching and learning, recommendations have little grounding. It is imperative that educational researchers continue to search and question tools being used to determine instructional practices that are most effective at promoting student learning.

Findings from this study provide useful information in exploring the complexities involved in observing effective teaching that promotes student learning. There are some factors that, in this study, have proven to make a difference in teacher effectiveness. There are other factors that have proven to make a difference when experience has been considered. There are, no doubt, other factors yet to be explored that may contribute as much or more to the equation of “best” instructional behaviors proven to increase student progress.

Teacher preparation programs are struggling with some of the same issues in trying to determine which practices should be taught in their colleges to provide the best training for future teachers. For the legitimacy of the profession and the strength of our future, we must take this matter seriously and rigorously determine what it is about teaching that makes a difference in achievement.

More than thirty years after Barr's study (1929), Wallen and Travers (1963) wrote of the impact of various teaching methods. They purported that when considering teacher effectiveness, various methods made little difference and one method could not be favored over another. A statistically significant difference in the means between those observed using the behavior and those not using the behavior was found in only three of the twenty-eight instructional behaviors reviewed in the observational component of this study. This study largely does this. It is time for a full reconceptualization of teaching and how it is connected to learning.

APPENDIX:
THE STUDY'S INSTRUMENTS

DEFINING TEACHER QUALITY
TEACHER SURVEY

CLASSROOM TECHNIQUES AND EXPERIENCES

What grade(s) do you currently teach? (Circle all that apply.)

(1) 3 4 5 6 7 8 9 10 11 12 Other (specify): _____

About how often, if at all, do your students engage in each of the following? (Circle one number on each line.)

0=never
1=not often (1 or 2 times per semester)
2=occasionally (1 or 2 times per month)
3=weekly (once a week)
4=often (a few times a week)
5=daily (every day or period)

		never	not often	occasionally	weekly	often	daily
(2)	Work individually on exercises, worksheets or workbooks	0	1	2	3	4	5
(3)	Work on a project that requires data collection and analysis	0	1	2	3	4	5
(4)	Review and discuss the work of other students	0	1	2	3	4	5
(5)	Work on group projects that extend for several days	0	1	2	3	4	5
(6)	Work on individual projects that take several days	0	1	2	3	4	5
(7)	Listen to teacher presentations	0	1	2	3	4	5
(8)	Explain their reasoning to the class	0	1	2	3	4	5
(9)	Discuss ideas for a sustained period	0	1	2	3	4	5
(10)	Answer factual questions in a whole class setting	0	1	2	3	4	5
(11)	Undertake service-learning projects	0	1	2	3	4	5
(12)	Reflect on their work and set future learning goals	0	1	2	3	4	5
(13)	Read from a textbook	0	1	2	3	4	5
(14)	Read a novel or primary source material	0	1	2	3	4	5
(15)	Debate their ideas in class	0	1	2	3	4	5
(16)	Use whiteboards	0	1	2	3	4	5
(17)	Use handouts	0	1	2	3	4	5

		never	not often	occasionally	weekly	often	daily
(18)	Are assessed by non-traditional means like journals or portfolios	0	1	2	3	4	5
(19)	Are assigned special long-term projects	0	1	2	3	4	5
(20)	Are tested using tests that the publishers include with textbooks/workbooks	0	1	2	3	4	5

About how often do you do each of the following in your classroom instruction? (Circle one number on each line.)

0=never
1=not often (1 or 2 times per semester)
2=occasionally (1 or 2 times per month)
3=weekly (once a week)
4=often (a few times a week)
5=daily (every day or period)

		never	not often	occasionally	weekly	often	daily
(21)	Introduce content through formal presentations	0	1	2	3	4	5
(22)	Arrange seating to facilitate student discussion	0	1	2	3	4	5
(23)	Use open-ended questions	0	1	2	3	4	5
(24)	Require students to explain their reasoning when giving an answer	0	1	2	3	4	5
(25)	Encourage students to communicate effectively	0	1	2	3	4	5
(26)	Encourage students to explore alternative methods for solutions	0	1	2	3	4	5
(27)	Allow students to work at their own pace	0	1	2	3	4	5
(28)	Embed assessment in regular class activities	0	1	2	3	4	5
(29)	Assign homework	0	1	2	3	4	5
(30)	Read and comment on the reflection students have written in their notebooks or journals	0	1	2	3	4	5

About how often do students in this class take part in each of the following types of activities as part of their instruction? (Circle one number on each line.)

0=never
1=not often (1 or 2 times per semester
2=occasionally (1 or 2 times per month)
3=weekly (once a week)
4=often (a few times a week)
5=daily (every day or period)

		never	not often	occasionally	weekly	often	daily
(31)	Participate in student-led discussions	0	1	2	3	4	5
(32)	Participate in discussions with the teacher to further understanding	0	1	2	3	4	5
(33)	Work in cooperative learning groups	0	1	2	3	4	5
(34)	Make formal presentations to the class	0	1	2	3	4	5
(35)	Read other non-textbook-related materials in class	0	1	2	3	4	5
(36)	Use concepts to interpret and solve problems	0	1	2	3	4	5
(37)	Work on solving real-world problems	0	1	2	3	4	5
(38)	Share ideas or solve problems with each other in small groups	0	1	2	3	4	5
(39)	Engage in hands-on activities	0	1	2	3	4	5
(40)	Play educational games	0	1	2	3	4	5
(41)	Work on models or simulations	0	1	2	3	4	5
(42)	Participate in field work	0	1	2	3	4	5
(43)	Write a description of a plan, procedure or problem-solving process	0	1	2	3	4	5
(44)	Write reflections in a notebook or journal	0	1	2	3	4	5
(45)	Use technology for learning or practicing skills	0	1	2	3	4	5
(46)	Use technology to develop conceptual understanding	0	1	2	3	4	5
(47)	Work on portfolios	0	1	2	3	4	5
(48)	Take short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank)	0	1	2	3	4	5
(49)	Take tests requiring open-ended responses (e.g., descriptions, justifications of solutions)	0	1	2	3	4	5
(50)	Engage in performance tasks for assessment purposes	0	1	2	3	4	5

Please indicate how well prepared you feel to do each of the following. (Circle one number on each line.)

		no preparation	not adequately prepared	somewhat prepared	fairly well prepared	well prepared	very well prepared
(51)	Lead a class of students using investigative strategies	0	1	2	3	4	5
(52)	Manage a class of students engaged in hands-on/project-based work	0	1	2	3	4	5
(53)	Help students take responsibility for their own learning	0	1	2	3	4	5
(54)	Recognize and respond to student diversity	0	1	2	3	4	5
(55)	Encourage students' interest in learning	0	1	2	3	4	5
(56)	Involve parents in the education of their children	0	1	2	3	4	5

Are you currently teaching in any subject area for which you feel inadequately prepared? (Mark an "X" in one box only.)

(57)

☐ I don't feel adequately prepared in: _____
(subject area)

☐ I feel well prepared in all areas I am teaching.

What percentage of your classroom time is devoted to learning? What percentage of time is spent managing student behavior and dealing with discipline?

(58) Time devoted to learning _____%

(59) Time devoted to managing student behavior/dealing with discipline _____%

How important for you is each of the following kinds of assessment in judging student learning? (Circle one number on each line.)

		not important	slightly important	moderately important	important	very important	extremely important
(60)	Multiple-choice tests	0	1	2	3	4	5
(61)	Essays tests	0	1	2	3	4	5
(62)	Portfolio of student work	0	1	2	3	4	5
(63)	Products of group projects	0	1	2	3	4	5
(64)	Standardized test results	0	1	2	3	4	5
(65)	Work samples	0	1	2	3	4	5
(66)	Performance tasks	0	1	2	3	4	5
(67)	Classroom participation	0	1	2	3	4	5
(68)	Other (specify: _____)	0	1	2	3	4	5

How much emphasis do you place on each of the following criteria in assessing student progress? (Circle one number on each line.)

The student showed increased ability to…		no emphasis	slight emphasis	minor emphasis	moderate emphasis	great emphasis	heavy emphasis
(69)	Recall factual information	0	1	2	3	4	5
(70)	Ask probing questions about subject matter	0	1	2	3	4	5
(71)	Apply what has been learned to new questions, situations, and subjects	0	1	2	3	4	5
(72)	Reflect on his progress	0	1	2	3	4	5

		no emphasis	slight emphasis	minor emphasis	moderate emphasis	great emphasis	heavy emphasis
(73)	Master basic skills	0	1	2	3	4	5
(74)	Express his own ideas about subject matter	0	1	2	3	4	5
(75)	Work with speed and accuracy	0	1	2	3	4	5
(76)	Provide constructive feedback to other students	0	1	2	3	4	5
(77)	Read for meaning	0	1	2	3	4	5

Using the scale below, please rate each of the following in terms of its IMPORTANCE for effective instruction in the grades you teach and how PREPARED you feel to do each one. (Circle one number in each section on each line.)

		Importance				Preparation			
		not important	somewhat important	fairly important	very important	not adequately prepared	somewhat prepared	fairly well prepared	very well prepared
(78)	Take students' prior understanding into account when planning curriculum and instruction	1	2	3	4	1	2	3	4
(80)	Make connections between disciplines	1	2	3	4	1	2	3	4
(82)	Have students work in cooperative learning groups	1	2	3	4	1	2	3	4
(84)	Have students participate in appropriate hands-on activities	1	2	3	4	1	2	3	4
(86)	Engage students in inquiry-oriented activities	1	2	3	4	1	2	3	4
(88)	Use technology	1	2	3	4	1	2	3	4
(90)	Use performance-based assessment	1	2	3	4	1	2	3	4
(92)	Use portfolios	1	2	3	4	1	2	3	4
(94)	Use informal questioning to assess student understanding	1	2	3	4	1	2	3	4
(90)	Use performance-based assessment	1	2	3	4	1	2	3	4
(92)	Use portfolios	1	2	3	4	1	2	3	4
(94)	Use informal questioning to assess student understanding	1	2	3	4	1	2	3	4

Please indicate the extent to which you agree or disagree with each of the following statements. (Circle one number on each line.)

		strongly disagree	disagree	disagree slightly	not sure	agree slightly	agree	strongly agree
TEACHING								
(96)	The attitudes and habits my students bring to class greatly reduce their chances for academic success.	1	2	3	4	5	6	7
(97)	By trying different teaching methods, I can significantly affect my students' achievement levels.	1	2	3	4	5	6	7
(98)	Many of the students I teach are not able to learn the material I am supposed to teach them.	1	2	3	4	5	6	7
(99)	My expectations for my students' learning are higher than they used to be.	1	2	3	4	5	6	7
(100)	There is really very little I can do to insure that most of my students achieve at a high level.	1	2	3	4	5	6	7
(101)	Students generally learn best in classes with students of similar abilities.	1	2	3	4	5	6	7
(102)	I am knowledgeable about current national standards in my content area(s).	1	2	3	4	5	6	7
(103)	I am willing to accept the noise that comes with an active classroom.	1	2	3	4	5	6	7
(104)	Encouraging student questions is more important than eliciting correct answers.	1	2	3	4	5	6	7
(105)	Students are less likely to misbehave when the activities and tasks are easy enough that everyone can do them successfully.	1	2	3	4	5	6	7
SCHOOL ENVIRONMENT								
(106)	I feel supported by colleagues to try out new ideas.	1	2	3	4	5	6	7
(107)	I have time during the regular school week to work with my peers on curriculum and instruction.	1	2	3	4	5	6	7
(108)	I have adequate access to materials/resources I need for my class.	1	2	3	4	5	6	7
(109)	I am provided adequate support from parents and the community.	1	2	3	4	5	6	7
(110)	I am provided the help I need in handling students who may be disruptive.	1	2	3	4	5	6	7

	strongly disagree	disagree	disagree slightly	not sure	agree slightly	agree	strongly agree
(111) Teachers should avoid grouping students by ability or level of performance.	1	2	3	4	5	6	7
(112) A lot of my ideas about teaching and learning come from my own experience as a student.	1	2	3	4	5	6	7
(113) Teachers should use the same standards in evaluating the work of all students in the class.	1	2	3	4	5	6	7
(114) It is impractical for teachers to tailor instruction to the unique interests and abilities of different students.	1	2	3	4	5	6	7
(115) When students work in groups, teachers can't really evaluate each individual's work.	1	2	3	4	5	6	7
(116) When working with students from low-income families, teachers should rely primarily on teacher-directed, focused, whole-group instruction.	1	2	3	4	5	6	7
(117) Teachers should pace their classes so that they cover the curriculum for their grade or course.	1	2	3	4	5	6	7
(118) The main job of the teacher is to teach subject matter.	1	2	3	4	5	6	7
(119) My views about teaching have changed substantially since I first began teaching.	1	2	3	4	5	6	7

Using the list below, please indicate the subject area(s), which you would consider your "current" teaching assignment(s). If you teach all subjects, please circle "01 self-contained classroom". (Circle all that apply).

(120) Self-contained classroom
(121) Reading
(122) English
(123) Language Arts
(124) History/Social Studies
(125) Any sciences
(126) Other math: _____
(127) Other subject:_____
(128) Elementary Mathematics
(129) Math Foundations
(130) Algebra I
(131) Algebra II
(132) Geometry

(133) I consider my main teaching assignment to be:_____
(list only one subject code)

During your <u>most recent typical week</u> of teaching, approximately how many hours did you spend <u>outside of your contracted hour</u> on each of the following types of activities? Write in number of hours. Do not leave any items blank. If no time is spent outside of your contracted hours, please write in "0" in the blank provided.

(134) Participating in activities related to governance, school site council or other school committees (e.g., attending meetings, planning) _____hours

(135) Being involved in one-on-one student activities (e.g., tutoring, counseling, advising) _____hours

(136) Sponsoring school-related activities involving student group interaction (e.g., sponsorship of extracurricular clubs or activities) _____hours

(137) Interacting with parents (e.g., conferences, phone calls) _____hours

(138) Attending school-related events and activities (e.g., planning a school dance, coordinating a holiday celebration) _____hours

(139) Mentoring or coaching other teachers _____hours

(140) Meeting with one or more colleagues to work on instruction _____hours

(141) Developing or reviewing curriculum or materials (e.g., textbook review, research for new courses) _____hours

(142) Developing standards or assessments for the school or district _____hours

(143) Grading homework, planning lessons, evaluating student work _____hours

During a typical full week of teaching, how much school time is OFFICIALLY SCHEDULED for you to plan and prepare?

(144)

_____hours _____minutes per week

☐ No time is officially scheduled for planning and preparation per week

Of the OFFICIALLY SCHEDULED school time for planning and preparation during a typical week, how much is ACTUALLY AVAILABLE to you for planning and preparation?

(145)

_____hours _____minutes per week

☐ No time is actually available for planning and preparation per week

During a typical full week of teaching, how much time do you think you NEED for planning and preparation?

(146)

_____hours _____minutes per week

☐ No time is needed for planning and preparation per week

During a typical full week of teaching, how much time do you ACTUALLY SPEND (total hours during and outside of school day) for planning and preparation?

(147)

_____hours _____minutes per week

☐ No time is actually spent for planning and preparation per week

Which category best describes the way YOUR classes at this school are organized? (Mark "X" only one box).

(148)

☐ Departmentalized instruction -You teach subject matter courses (e.g., biology, history) to several classes of different students all or most of the day.

☐ Elementary class – You teach only one subject (e.g., math, English, Language Arts, Reading) in an elementary school.

☐ Self-contained class – You teach multiple subjects to the same class of students all or most of the day.

☐ Team teaching – You collaborate with one or more teachers in teaching multiple subjects to the same class of students.

☐ "Pull Out" Class – You provide instruction (e.g., special education, reading) to certain students who are released from their regular classes.

☐ Other: (specify) _____

During your most recent FULL WEEK of teaching, approximately how many hours did you spend teaching each of these subjects at this school? Report hours to the nearest whole hour, do not record fractions or minutes. If you did not teach a particular subject during the week, mark the "none" box. If you are not assigned to teach a particular subject, mark the "NA" box for not applicable. (Mark an "X" or number of hours on each line.)

(149)	English	_____hours per week	____none	_____NA
(150)	Reading	_____hours per week	____none	_____NA
(151)	Language Arts	_____hours per week	____none	_____NA
(152)	Mathematics	_____hours per week	____none	_____NA
(153)	Social Studies/History	____hours per week	____none	_____NA
(154)	Science	_____hours per week	____none	_____NA

What do you consider to be your greatest strengths as a teacher? Think about both areas of content mastery and instructional strategies when answering this question. (Please print clearly.)

(155) Strengths:

What areas of your teaching do you think need improvement? Think about both areas of content mastery and instructional strategies when answering the question. (Please print clearly.)

(156) Areas that need improvement:

During the last 12 months, have you been involved in any of the following activities related to your teaching? (Circle "yes" or "no" on each line.)

(157)	Served as department chair/grade level chair/team leader	yes	no
(158)	Developed or piloted new curricula	yes	no
(159)	Held a leadership position in a state or national professional organization	yes	no
(160)	Formally mentored beginning teacher(s)	yes	no
(161)	Supervised student teacher(s)	yes	no
(162)	Conducted in-services or workshops for teachers	yes	no
(163)	Made observational visits to other schools	yes	no
(164)	Made presentations to non-teaching groups (e.g. school board, parents, community groups)	yes	no
(165)	Conducted individual or collaborative research on a topic of interest to you	yes	no
(166)	Represented the school or district on an instructional reform/initiative project	yes	no
(167)	Other (specify: _____)	yes	no

This question concerns how teachers interact with each other in your school. Please indicate the frequency with which YOU do each of the following. (Circle one number on each line.)

0=never
1=not often (1 or 2 times per semester)
2=occasionally (1 or 2 times per month)
3=weekly (once a week)
4=often (a few times a week)
5=daily (every day or period)

		never	not often	occasionally	weekly	often	daily
(168)	Share ideas on teaching with other teachers	0	1	2	3	4	5
(169)	Observe another teacher teaching	0	1	2	3	4	5
(170)	Be observed by another teacher	0	1	2	3	4	5
(171)	Teach with a colleague	0	1	2	3	4	5
(172)	Discuss with other teachers what you/they have learned at a workshop or conference	0	1	2	3	4	5
(173)	Analyze student work with other teachers	0	1	2	3	4	5
(174)	Discuss particular lessons that were not very successful	0	1	2	3	4	5
(175)	Discuss beliefs about teaching and learning	0	1	2	3	4	5

		never	not often	occasionally	weekly	often	daily
(176)	Discuss how to help students having problems	0	1	2	3	4	5
(177)	Discuss common challenges in the classroom	0	1	2	3	4	5
(178)	Work together to develop teaching materials or activities for particular classes	0	1	2	3	4	5
(179)	Discuss and attempt to solve program- or school-level problems	0	1	2	3	4	5
(180)	Plan a curriculum with other teachers	0	1	2	3	4	5

To what extent, if any, do you agree with the following statements. (Circle one number on each line.)

		strongly disagree	disagree	disagree slightly	not sure	agree slightly	agree	strongly agree
(181)	My teaching and learning activities are focused on helping students achieve the standards.	1	2	3	4	5	6	7
(182)	My curriculum and instructional strategies emphasize higher-level thinking and problem-solving skills.	1	2	3	4	5	6	7
(183)	I use a variety of teaching strategies and learning activities to help students learn.	1	2	3	4	5	6	7
(184)	My instruction includes the active participation of students.	1	2	3	4	5	6	7
(185)	I teach students how to assess their own progress and how to set their own learning goals so they may become independent learners.	1	2	3	4	5	6	7
(186)	I give students a variety of ways to show what they have learned (for example, projects, portfolios, presentations).	1	2	3	4	5	6	7
(187)	I give students feedback on student progress and provide suggestions.	1	2	3	4	5	6	7
(188)	I give parents feedback on student progress and provide suggestions.	1	2	3	4	5	6	7
(189)	I feel safe at school.	1	2	3	4	5	6	7
(190)	Most of the students in our school are well behaved.	1	2	3	4	5	6	7
(191)	Discipline problems at our school are handled quickly and fairly.	1	2	3	4	5	6	7
(192)	I can freely express my opinions or concerns to the administrator(s).	1	2	3	4	5	6	7

		strongly disagree	disagree	disagree slightly	not sure	agree slightly	agree	strongly agree
(193)	I encourage and welcome parents to come to my classroom.	1	2	3	4	5	6	7
(194)	At the beginning of the school year or semester, I clearly inform students about what they are expected to know and be able to do.	1	2	3	4	5	6	7
(195)	At the beginning of the school year or semester, I clearly inform parents about what students are expected to know and be able to do.	1	2	3	4	5	6	7
(196)	I am satisfied with how well my students are achieving the standards.	1	2	3	4	5	6	7
(197)	I am satisfied with my school's leadership.	1	2	3	4	5	6	7
(198)	I am teaching in my area of certification.	1	2	3	4	5	6	7
(199)	I am confident about my ability to teach.	1	2	3	4	5	6	7
(200)	I don't think I am a very good teacher right now.	1	2	3	4	5	6	7
(201)	It would be nice if others could or would come into my classroom and demonstrate teaching strategies.	1	2	3	4	5	6	7
(202)	I have few discipline problems with my students.	1	2	3	4	5	6	7
(203)	I usually look forward to coming to school to teach.	1	2	3	4	5	6	7

Additional comments on **CLASSROOM TECHNIQUES AND EXPERIENCES** (Please print clearly.):

PROFESSIONAL DEVELOPMENT

PLEASE CONSIDER LAST SCHOOL YEAR'S EXPERIENCES WHEN ANSWERING THE FOLLOWING QUESTIONS ON PROFESSIONAL DEVELOPMENT.

Please indicate the extent to which professional development activities (district, site-based or personal) and support have made the following contributions to you as a teacher. (Circle one number on each line for each section.)

"District" means offered and/or required by the district.
"Site-based" means offered at a specific site/school based on individual school need.
"Personal" means self-selected, not offered or required by individual school or district (i.e., reading, conferences, college courses, etc.).

1 = not at all
2 = a little
3 = somewhat
4 = a lot

	DISTRICT				SITE-BASED				PERSONAL			
(204) Deepened my grasp of the subject matter I taught	1	2	3	4	1	2	3	4	1	2	3	4
(207) Increased my knowledge beyond basic instruction and assessment	1	2	3	4	1	2	3	4	1	2	3	4
(210) Improved my skills to meet instructional needs of the student population at this school (e.g., English language learners or students from diverse cultural backgrounds)	1	2	3	4	1	2	3	4	1	2	3	4
(213) Improved my classroom management, allowing me to try new instructional activities	1	2	3	4	1	2	3	4	1	2	3	4
(216) Increased my confidence and responsiveness in interactions with parents	1	2	3	4	1	2	3	4	1	2	3	4
(219) Improved my ability to consistently identify instructional goals appropriate to the subject matter I taught	1	2	3	4	1	2	3	4	1	2	3	4
(222) Increased my effectiveness at promoting student learning	1	2	3	4	1	2	3	4	1	2	3	4
(225) Helped me ask for additional assistance and feedback when I needed it	1	2	3	4	1	2	3	4	1	2	3	4
(228) Helped me understand the way my school/district and its administration worked	1	2	3	4	1	2	3	4	1	2	3	4

1 = not at all
2 = a little
3 = somewhat
4 = a lot

	DISTRICT				SITE-BASED				PERSONAL			
(231) Provided information that was new to me	1	2	3	4	1	2	3	4	1	2	3	4
(234) Caused me to seek further information or training	1	2	3	4	1	2	3	4	1	2	3	4
(237) Was generally a waste of my time	1	2	3	4	1	2	3	4	1	2	3	4
(240) Increased my use of inquiry or action research	1	2	3	4	1	2	3	4	1	2	3	4
(243) Improved my ability to teach special needs students	1	2	3	4	1	2	3	4	1	2	3	4
(246) Improved my ability to plan instruction	1	2	3	4	1	2	3	4	1	2	3	4
(249) Increased my understanding of the applications of technology in everyday life	1	2	3	4	1	2	3	4	1	2	3	4
(252) Made me familiar with new materials and equipment that I can use in my teaching.	1	2	3	4	1	2	3	4	1	2	3	4
(255) Taught me about innovative ways to use standard materials and equipment in my field	1	2	3	4	1	2	3	4	1	2	3	4
(258) Increased my knowledge of current educational issues	1	2	3	4	1	2	3	4	1	2	3	4
(261) Increased my ability to use data to inform instruction	1	2	3	4	1	2	3	4	1	2	3	4

To what degree, if any, do you agree that the following would strengthen the district professional development program? (Circle one answer on each line.)

	strongly disagree	disagree	disagree slightly	not sure	agree slightly	agree	strongly agree
(264) Cover new topics (specify_____)	1	2	3	4	5	6	7
(265) Work on team building within schools	1	2	3	4	5	6	7
(266) Have more mandatory training	1	2	3	4	5	6	7
(267) Offer more follow-up after initial training	1	2	3	4	5	6	7
(268) Have more choices for training	1	2	3	4	5	6	7
(269) Use presenters from the local school district	1	2	3	4	5	6	7
(270) Other (specify)_____	1	2	3	4	5	6	7

Have you participated in any professional development activities that focused on the following topics? If yes, to what degree, if any, did this activity benefit your instruction? (Circle one response in each section on each line.)

	Participated?		Not at all	Slight degree	Moderate degree	Good degree	Great degree	Not applicable
(271) In-depth study of your main subject area (i.e., reading or mathematics)	Yes	No	1	2	3	4	5	NA
(273) Methods of teaching your main subject area	Yes	No	1	2	3	4	5	NA
(275) Applications of technology to education	Yes	No	1	2	3	4	5	NA
(277) Other related areas (specify) _____	Yes	No	1	2	3	4	5	NA

Consider all of the professional development activities in which you participated. Please indicate the extent to which you agree with the following statements. (Circle one number on each line.)

In general, professional development available to me…	strongly disagree	disagree	disagree slightly	not sure	agree slightly	agree	strongly agree
(279) Recognizes and builds on individual teacher's knowledge and experience.	1	2	3	4	5	6	7
(280) Promotes collaboration and joint work among teachers	1	2	3	4	5	6	7
(281) Is sustained over time, with ample participant follow-up and teacher support	1	2	3	4	5	6	7
(282) Is a series of single events with little or no follow-up	1	2	3	4	5	6	7
(283) Focuses on subject matter content and how to teach it	1	2	3	4	5	6	7
(284) Is a good fit with what I need or want in my current teaching assignment(s)	1	2	3	4	5	6	7
(285) Is developed and organized by teachers in my school	1	2	3	4	5	6	7
(286) Focuses on how teaching and learning build on students' backgrounds and experiences	1	2	3	4	5	6	7

In general, to what degree are the professional development opportunities tailored to meet your needs? (Circle only one response.)

	not at all	slight degree	moderate degree	good degree	great degree
(287) Tailored to meet my needs	1	2	3	4	5

During the last 12 months, how FREQUENTLY have you participated in the following activities related to teaching? For any activity in which you participated, indicate the extent you believe the activity has IMPROVED YOUR CLASSROOM TEACHING. Include any professional development activities you participated in, but exclude any activities you participated in during preservice training. (Circle one response in EACH section on each line.)

	Frequency of Activity						Improved Teaching				
	never	a few times a year	once a month	2 to 3 times a month	at least once a week		not a lot	somewhat	moderate	a lot	not applicable
(288) Common planning period for teachers	0	1	2	3	4	5	1	2	3	4	NA
(290) Being mentored by another teacher in a formal relationship	0	1	2	3	4	5	1	2	3	4	NA
(292) Mentoring another teacher in a	0	1	2	3	4	5	1	2	3	4	NA
(294) Networking with teachers outside your school	0	1	2	3	4	5	1	2	3	4	NA
(296) Regularly scheduled collaboration with other teachers, excluding meetings held for administrative purposes	0	1	2	3	4	5	1	2	3	4	NA
(298) Other (please describe)_____	0	1	2	3	4	5	1	2	3	4	NA

Additional comments on **PROFESSIONAL DEVELOPMENT** (Please print clearly.):

CERTFICATIONS AND ENDORSEMENTS

Do you have a teaching certificate in the state in which you are currently teaching? (Mark an "X" in one box only.)

(300)

- ☐ Yes
- ☐ No

What type of teaching certificate do you hold? (Mark an "X" in one box only.)

(301)

- ☐ Regular or standard state certificate, or advanced professional certificate
- ☐ Provisional or other type of certificate given to persons who are still participating in what the state calls an "alternative certification program"
- ☐ Probationary certificate (the initial certificate issued after satisfying all requirements except the completion of a probationary period)
- ☐ Temporary certificate (requires some additional college coursework and/or student teaching before regular certification can be obtained)
- ☐ Emergency certificate or waiver (issued to persons with insufficient teacher preparation who must complete a regular certification program in order to continue teaching)
- ☐ Other (please specify: _____)
- ☐ Uncertain

Are you currently teaching in any subject area for which you are NOT professionally certified? (Mark an "X" in one box only.)

(302)

- ☐ I am not professionally certified in: _____
- ☐ I am professionally certified in all areas I am teaching.

How did you earn your regular or standard state certificate or advanced professional certificate in your MAIN teaching assignment field? (Mark "X" in one box only.)

(303)

- ☐ As part of a bachelor's degree program
- ☐ As part of a "5th year" program
- ☐ As part of a master's degree program
- ☐ After I began teaching, as part of an alternative program
- ☐ Before I began teaching, as part of an alternative program
- ☐ Through continuing professional development
- ☐ Other (Please specify: _____)

Additional comments on **CERTIFICATIONS AND ENDORSEMENTS** (Please print clearly.):

COLLEGE PREPARATION AND EXPERIENCE

Please check the box next to the degree(s) you have earned. From the given fields, list your major and minor fields of study CODES for each degree. If you do not have a second major or minor, please write "none".

***for example*: Major field is Education with a minor in Mathematics.**
Major field code - <u>01</u> minor field code - <u>11</u>

Use CODES given on last page of this survey.

	Major Field Code	Minor Field Code
❑ Bachelor's Degree (304-05)	_____	_____
❑ Master's Degree (306-07)	_____	_____
❑ Doctorate Degree (308-09)	_____	_____

❑ Other degree (specify:_____) (310)

University granting your bachelor's degree:

(311) Name of institution: _____

(312) City, State: _____

University granting your master's degree:

(313) Name of institution: _____

(314) City, State: _____

Other degree granting institutions:

(315) Degree: _____

(316) Name of institution: _____

(317) City, State: _____

Are you currently working toward an advanced degree? If yes, please indicate the degree you are pursuing, the institution and the area in which you are specializing. (Mark an "X" in one box only.)

(318)

☐ Yes Degree: _____

Institution: _____

Specialization: _____

☐ No

In order to get a general sense of your educational background, please circle the titles of courses you have taken for credit as part of your postsecondary education. Do not be concerned about whether the titles match the classes you took, or if you do not have total recall of this information. (Circle all that apply.)

EDUCATION COURSES

(319) Computers/technology in the classroom
(320) Curriculum and curriculum theory
(321) Education administration
(322) Education/instruction media design
(323) Educational assessment, testing and measurement
(324) Educational psychology
(325) Educational statistics
(326) Evaluation and research in education
(327) Instruction, methods, and materials
(328) Mathematics in education
(329) School psychology
(330) Science teacher education
(331) Social/historical/philosophical foundations of education
(332) Teacher education: intermediate or secondary school
(333) Other (specify: _____)

Did your preparation for teaching include… (Circle yes or no on each line.)

(334) coursework in how to select and adapt instructional materials?	yes	no
(335) coursework in learning theory or psychology appropriate to the age of the students you teach?	yes	no
(336) your observation of other classroom teaching?	yes	no
(337) feedback on your teaching?	yes	no

(338) practice [student] teaching? (Mark an "X" in one box only.)

- ☐ I had no practice teaching
- ☐ 4 weeks or less
- ☐ 5-9 weeks
- ☐ 10 weeks or more

Additional comments on **COLLEGE EDUCATION AND EXPERIENCE** (Please print clearly.):

PERSONAL BACKGROUND

Your gender? (Mark an "X" in one box only.)

(339)

- ☐ Female
- ☐ Male

How many children do you have? (Mark an "X" in one box only.)

(340)

- ☐ 0 children
- ☐ 1 child
- ☐ 2 children
- ☐ 3-5 children
- ☐ 6 or more children

How many years separate your high school graduation and your undergraduate college admission? (Mark an "X" in one box only).

(341)

- ☐ less than 1 year
- ☐ 1-3 years
- ☐ 3-5 years
- ☐ more than 5 years

During a typical full week of teaching, approximately how much time do you spend on each of the following activities OUTSIDE your job as a teacher? If no hours are spent, mark an "X" in the "no time" box.

(342) Working at a job in the education field, outside of teaching

_____hours _____minutes

☐ No time/do not work other job than teaching

(343) Working at an occupation outside of education

_____hours_____minutes

☐ No time/do not work other job than teaching

(344) Attending a college or university

_____hours _____minutes

☐ No time/do not work other job than teaching

Previous areas of employment, another occupation full time for at least 1 year prior to teaching. (Mark an "X" for all that apply.)

(345-357)

- ☐ Have held no other jobs for at least 1 year prior to teaching
- ☐ Unskilled (fast-food worker, etc.)
- ☐ Semi-skilled white collar (bank teller, secretary)
- ☐ Semi-skilled blue collar (assembly line)
- ☐ Skilled (certified trade, electrician, plumber)

- ☐ Owner or operator of a small business
- ☐ College or university professor
- ☐ White collar public service (social worker, childcare, minister)
- ☐ Blue collar public service (police officer, firefighter)
- ☐ Manager (public/private corporation, administrator)
- ☐ Other professional (dentist, lawyer)
- ☐ Military
- ☐ Other (specify): _____

As of June 2002 (end of last school year), how many years will you have been… (Write the number of years in the space provided. If less than 1 year, indicate this with a "0".)

(358) a full-time teacher? _____years

(359) a teacher in this district? _____years

(360) a teacher in this school? _____years

(361) a teacher in this grade? _____years

(362) a teacher in this content area? _____years

How much time do you feel satisfied with your job? (Mark an "X" in one box only.)

(363)

- ☐ All of the time
- ☐ Most of the time
- ☐ Some of the time
- ☐ Almost never

Upon entering undergraduate school, what was your first declared major?
for example: Major field code - <u>01</u> minor field code – <u>11</u>
The major field is Education with a minor in Mathematics.

Use CODES given on last page of this survey.

(364) 1st declared major code: _____

Which of the following settings best describes the area where you spent the majority of your pre-college life? (Mark an "X" in one box only.)

(365)

- ☐ Small town/rural (population less than 25,000)
- ☐ Town (population 25,000 to 100,000)
- ☐ City (population 100,000 to 500,000)
- ☐ Urban (population more than 500,000)
- ☐ Overseas (specify where): _____

When did you decide to become a teacher? (Mark an "X" in one box only.)

(366)

- ☐ Always wanted to be a teacher
- ☐ Before elementary school
- ☐ Elementary/primary school
- ☐ Middle/junior high school
- ☐ High school
- ☐ College
- ☐ Other (specify): _____

Have you ever worked with young people in any of the following ways. (Mark an "X" for all that apply.)

(367-376)

- ☐ Religious-school teacher or aide
- ☐ Camp counselor
- ☐ Teacher's aide
- ☐ Preschool aide
- ☐ Tutoring (including piano, etc.)
- ☐ Sports coaching
- ☐ Babysitting
- ☐ Parent
- ☐ Other (specify): _____
- ☐ I have no experience working with young people, outside of the classroom.

For each of the following kinds of professional organizations, please indicate your level of involvement. If you hold a membership, please specify the professional organization. (Circle one number on each line.)

	not a member	member, no involvement	member, somewhat involved	active member	active in leadership role
(377) Subject-area association (specify: _____)	1	2	3	4	5
(378) National teaching association (specify: _____)	1	2	3	4	5
(379) State or regional teaching association (specify: _____)	1	2	3	4	5
(380) Local teacher association (specify: _____)	1	2	3	4	5
(381) On-line teacher network (specify: _____)	1	2	3	4	5
(382) Other teacher network (specify: _____)	1	2	3	4	5

Educational backgrounds of parents/guardians. (Circle one for mother/guardian and one for father/guardian.)

(383) ***Mother/guardian***	(384) ***Father/guardian***
Less than high school degree	Less than high school degree
High school graduate or GED certificate	High school graduate or GED certificate
Some college	Some college
College degree	College degree
Post-graduate or professional degree	Post-graduate or professional degree
Other (specify: _____)	Other (specify: _____)

Which of the following best described your parent/guardians' occupations <u>at the time you were in high school?</u>

(385) Father/guardian's occupation (Choose one from the list below): _____

(386) Mother/guardian's occupation (Choose one from the list below): _____

Not employed
Unskilled (fast-food worker)
Semi-skilled white collar (bank teller, secretary)
Semi-skilled blue collar (assembly line)
Skilled (certified trade, electrician, plumber)
Teacher, college or university professor
White collar public service (social worker, minister)
Blue collar public service (police office, firefighter)
Manager (public/private corporation, administrator)
Other professional (dentist, lawyer)
Other _____

USE THIS CHART FOR QUESTIONS ON PAGE 20AND 25

Education

01 All education fields (including general education, subject area education, and special education, leadership and administration)

Science

02 Biology/life science
03 Chemistry
04 Geology/earth sciences
05 Physics
06 Other science-Please specify above

Humanities

07 English/literature/classics
08 Foreign language
09 Philosophy, religion, theology
10 Other humanities (specify)

Mathematics

11 Mathematics
12 Statistics

Social Science

13 History
14 Political Science
15 Economics
16 Psychology
17 Sociology
18 Other social science (specify)

General fields

19 Art, fine and applied
20 Business and management
21 Communications/journalism
22 Computer/information science
23 Dance, drama, theater, or music
24 Engineering
25 Health professions and occupations
26 Physical education

27 Interdisciplinary studies
28 All others (specify)

OBSERVATION INDICATOR CHECKLIST

OBSERVATION INDICATOR CHECKLIST

CHECKLIST 2: TO BE COMPLETED BEFORE SOA

1. Planning

a.	1. _____	The teacher demonstrated an appropriate sequence of ideas.
	2. _____	The teacher provided the goal of the lesson.
b.	3. _____	The teacher's transitions were smooth/seamless.
	4. _____	The teacher had materials readily accessible.
	5. _____	The teacher's lesson showed organization.
	6. _____	The teacher managed a class of students engaged in hands-on/project-based work.
c.	7. _____	The teacher asked questions about a homework assignment.
	8. _____	The teacher utilized graphic organizers such as K-W-L.
	9. _____	The teacher used the board as well as speaking.
	10. _____	The teacher connected learning to previous lessons/units.
	11. _____	The teacher might have commented, "From looking over your tests…"
	12. _____	The teacher identified/addressed prior conceptions and/or misconceptions.
	13. _____	The teacher assigned homework.
	14. _____	There was evidence of the teacher's awareness of students'
		_____ level of preparedness
		_____ prior knowledge
		_____ learning styles
	15. _____	The teacher used a variety of teaching strategies and learning activities to help students learn. *(also 1e)*
d.	16. _____	The teacher used a video clip (or other resource) that explained the concept…
	17. _____	The teacher used manipulatives that were useful in presenting the ideas the lesson.
	18. _____	The teacher's text supported ideas presented by him/her.
	19. _____	The teacher used examples that were appropriate to the group (current/fad resources used, 8-track tapes).
	20. _____	The students used (individual) whiteboards.
	21. _____	The students read from a textbook (silently or aloud).
	22. _____	The students read other non-textbook-related materials in class.

23. _____ The students read a novel or primary source material.

24. _____ The students used technology for learning or practicing skills.

25. _____ The students used handouts.

26. _____ The students worked on models or simulations.

27. _____ The students worked on:

_____ exercises

_____ worksheets

_____ workbooks

e. 28. _____ The teacher called on many different children.

29. _____ The students were in diverse groups.

30. _____ The teacher used stories/manipulatives that were culturally relative to content/ideas/students.

31. _____ The teacher used examples that reflected age and culture of students.

32. _____ The teacher applied knowledge of how the students learn and develop to create developmentally appropriate learning opportunities.

33. _____ The seating arrangement reflected equity and diversity.

34. _____ The teacher used a variety of teaching strategies and learning activities to help students learn. *(also 1c)*

35. _____ The students played educational games.

36. _____ The students engaged in hands-on activities.

f. 37. _____ The teacher arranged seating to facilitate student discussion.

38. _____ The students were working together.

39. _____ The students turned to a partner for a short conversation.

40. _____ The teacher's instruction included the active participation of students.

41. _____ The students worked in cooperative learning groups.

42. _____ The students shared ideas or solved problems with each other

_____ in small groups

_____ in pairs

2. Implementation

a. 1. _____ The teacher introduced content through formal presentations.

2._____ The teacher's goal was to teach the basic facts of a lesson and the teacher used an appropriate strategy(s) to reach this goal.

	3. _____	The students were required to explain why they did something. (The teacher asked students to explain their reasoning.)
	4. _____	The students worked on projects that required data collection and/or analysis.
b.	5. _____	The teacher used multiple examples (teacher).
	6. _____	The teacher continued to explain a concept the same way when students did not understand the first time. *Doing it the same way when students don't understand the first time.* (Reverse coded).
	7. _____	The teacher was not tied to a textbook or notes (free to walk around or look up at students).
	8. _____	The teacher demonstrated the ability to ask appropriate spontaneous questions.
	9. _____	The teacher was comfortable to say she did not know the answer but would help student find it.
	10. _____	The teacher's voice tone suggested confidence.
c.	11. _____	The teacher demonstrated appropriate wait time with questioning.
	12. _____	The teacher used reteaching when needed.
	13. _____	The teacher allowed appropriate wait time for the given task.
	14. _____	The teacher allowed students to work at their own pace.
	15. _____	Students remained engaged in the lesson.
d.	16. _____	The teacher used/asked open-ended questions.
	17. _____	The students were asked to make judgments.
	18. _____	The students were required to synthesize.
	19. _____	The students were asked questions that required application of knowledge.
	20. _____	The teacher's curriculum and instructional strategies emphasized higher level thinking and problem-solving skills.
	21. _____	The teacher asked higher order questions.
	22. _____	The students explained their reasoning to the class.
e.	23. _____	The teacher demonstrated appropriate wait time with questioning.
f.	24. _____	The students shared out with partners, small or whole group.
	25. _____	The students wrote reflections in a notebook or journal.
	26. _____	The students reflected on their work and/or set future learning goals.
	27. _____	The students wrote a description of a plan, procedure or problem solving process.

3. Content

a.	1. _____	The teacher demonstrated knowledge of the content/discipline applicable to students' level/s.
	2. _____	The teacher used students' prior knowledge.
	3. _____	The teacher strengthened prior knowledge with new ideas.
	4. _____	The students were actively engaged.
	5. _____	The students listened to teacher presentations.
b.	6. _____	The teacher presented the subject matter in multiple ways as necessary.
	7. _____	The teacher lead a class of students using investigative strategies.
	8. _____	The teacher responded appropriately with questions or responses to help the students understand the concept.
	9. _____	The students participated in discussions with the teacher to further understanding.
	10. _____	The students answered factual questions in a whole class setting.
c.	11. _____	The students were on-task.
	12. _____	The students asked appropriate questions.
	13. _____	The students made formal presentations to the class.
	14. _____	The students debated their ideas in class.
	15. _____	The students responded appropriately to teacher questions.
d.	16. _____	The teacher used phrases and words defined by the subject being discussed.
	17. _____	The teacher encouraged the appropriate vocabulary from students.
e.	18. _____	The teacher used examples to which the students could relate.
	19. _____	The teacher used examples current to trends and activities of students.
	20. _____	The students worked on solving real-world problems.
	21. _____	The teacher made connections to other disciplines/content areas.

4. Classroom culture

a.	1. _____	The teacher encouraged students to communicate effectively.
	2. _____	The teacher gave verbal comments to students.
	3. _____	The teacher called on many students.
	4. _____	The teacher gave ample think/wait time.
	5. _____	The teachers might have read what students had written and

positively responded to it verbally.

6. _____ The students felt comfortable to interject questions.
7. _____ The class rules stated something about respect.
8. _____ The students/teacher did not interrupt each other.
9. _____ The students freely communicated.
10. _____ The teacher walked around responding to group needs, asking questions, inquiring, etc.
11. _____ The teacher accepted the noise that came with an active classroom.
12. _____ The teacher displayed evidence of coaching/encouraging.

b.

13. _____ The students were comfortable to respond to each other.
14. _____ The students freely communicated.
15. _____ The students shared ideas or solved problems with each other

    _____ in small groups

    _____ in pairs
16. _____ The students responded to group needs, asking questions, inquiring, etc.
17. _____ The students shared comments/responses and used good listening skills as others shared comments/responses.
18. _____ The students participated in student/teacher-led discussions.
19. _____ The students worked in cooperative learning groups.

c.

20. _____ The teacher's questions challenged students to think beyond "one" answer.
21. _____ The climate was such that constructive criticism was accepted as necessary to enhance the learning process.
22. _____ The teacher required students to explain their reasoning when giving an answer.
23. _____ The teacher encouraged questions as much or more than answers.
24. _____ The teacher encouraged students to explore alternative methods for solutions.
25. _____ The students wrote a description of a plan, procedure or problem solving process.
26. _____ The students explained their reasoning to the class.
27. _____ The students discussed ideas for a sustained period.
28. _____ The teacher encouraged the generation of student ideas, questions and/or contributions.

5. Assessment

a. 1._____ The students were taught how to assess their own progress and how to set their own learning goals in order to become independent learners.

2. _____ The students checked their own work.

b. 3. _____ The work of other students was reviewed/discussed.

c. 4. _____ There was evidence of:

_____ multiple-choice tests being given.

_____ work samples (any work) being collected or reviewed.

_____ questions & answers (Q & A)

_____ standardized tests being administered and/or results discussed.

_____ essay tests being given.

5. _____ The teacher gave students feedback on their progress and provided suggestions.

6. _____ The students were tested using tests that the publishers include with textbooks/workbooks.

d. 7. _____ There was evidence of

_____ student portfolios.

_____ special long-term projects.

_____ performance tasks (product or performance that demonstrated a skill).

_____ products of group projects.

8. _____ The teacher provided students a variety of ways to show what they had learned (for example: projects, portfolios, presentations).

e. 9. _____ The teacher assessed by walking around.

6. Behavioral Management

a. 1. _____ The teacher managed behavior using proximity.

2. _____ The teacher managed behavior with students individually.

3. _____ The teacher addressed misbehavior publicly.

4. _____ The teacher had few/no discipline problems.

b. 5. _____ The teacher had few/no discipline problems.

6. _____ The teacher managed behavior using proximity.

7. _____ The teacher managed behavior with students individually.

8. _____ The teacher addressed misbehavior publicly.

9. _____ The teacher allowed misbehavior to disrupt the learning environment.

c. 10. _____ Rules/classroom procedures were visibly posted in classroom.

d. 11. _____ minutes Time spent devoted to instructional activities

e. 12. _____ minutes Time spent devoted to managing student behavior/dealing with discipline

ENVIRONMENTAL CHECKLIST

ENVIRONMENTAL CHECKLIST

Condition of Room *Comments*	*Condition*		
Windows	good	need repair	none
Walls	good	need repair	none
Ceiling	good	need repair	none
Lights	good	need repair	none
Desks	good	need repair	none
Chairs	good	need repair	none
Tables	good	need repair	none
Bulletin boards	good	need repair	none
Books	good	need repair	none
Bookshelves	good	need repair	none
Library area/centers	good	need repair	none
Computers/Printers/Tape players	good	need repair	none
Pictures, Graphs, Charts on walls	good	need repair	none
Other specify_____	good	need repair	none

Contents of Room

_____ Resources good needs repair

Number of books _____

Other (specify)_____

_____ Learning centers good needs repair

Number of learning centers _____

Audio equipment _____

Other materials in learning centers _____

_____ Technology good needs repair

Number of computers in classroom _____

Number of printer(s) _____

Other technology

Specify_____

_____ Evidence of student work

Describe:

_____ Evidence of *posted* behavioral expectations

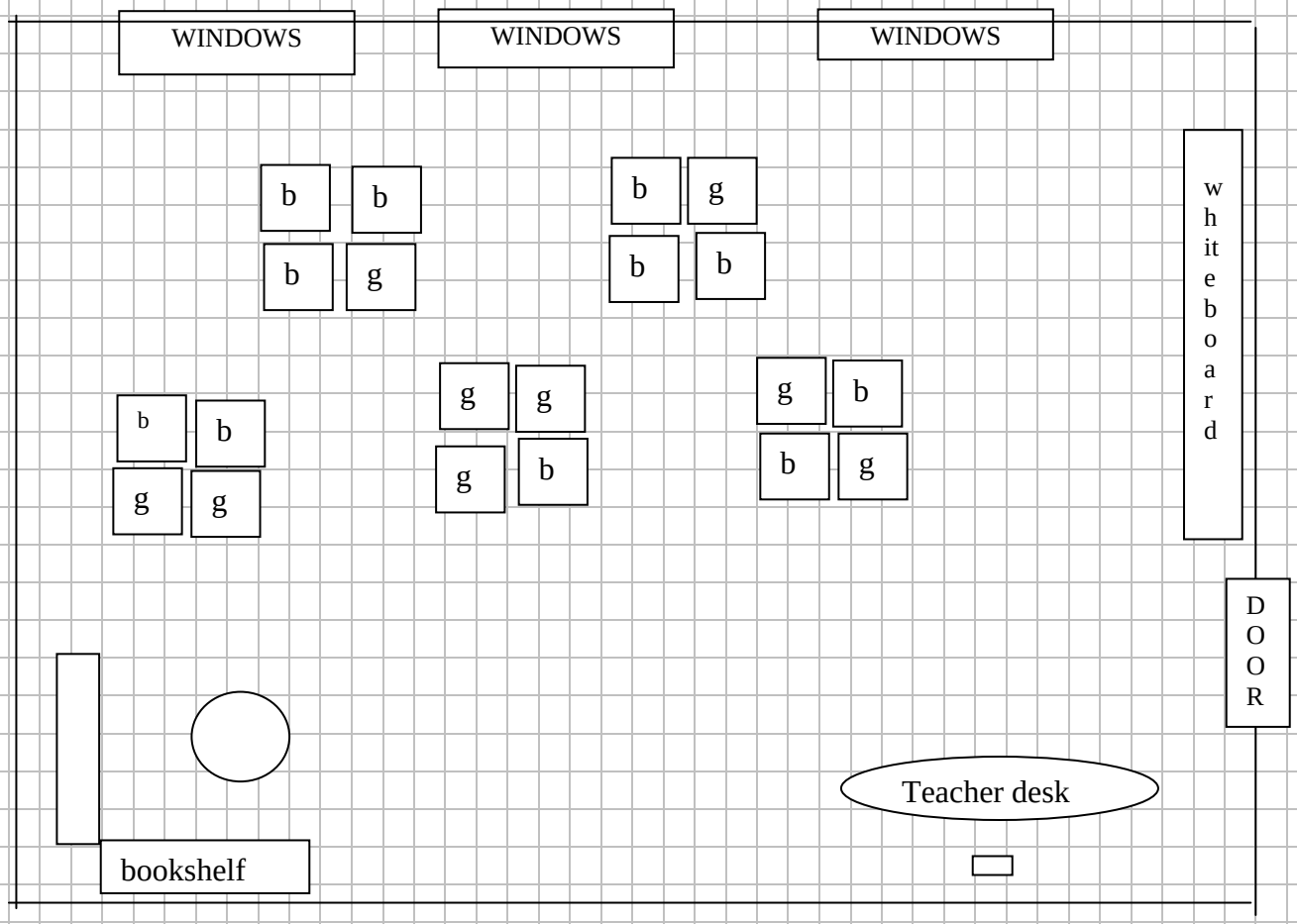
Describe:

Seating Arrangements (Check all that apply)

- ☐ Assigned seats
- ☐ Random seating
- ☐ Desks in rows and columns
- ☐ Desks in semi-circles
- ☐ Desks in clusters
- ☐ Tables not desks
- ☐ Whole group work/small group work
- ☐ Other specify_____

Narrative of classroom contents, conditions and other:

EXAMPLE OF CLASSROOM SKETCH AND NARRATIVE



NARRATIVE SUMMARY:
Craft a paragraph that provides a description of the room as a context for instruction within the particular intervention you are observing. Try to provide the analyst a brief glimpse of features of the classroom environment that seem prominent or important but may not be well captured elsewhere in this section. The interest in is describing what kind of environment this is for learning. For example you might highlight things like: quantity of student work on the walls, whether there are multiple forms of reading material and evidence of students' writing, if most displays in the classroom are of student work. Often rooms "feel" inviting, barren, cramped, rundown, chaotic, filled by student products, dominated by teacher products or commercial material, and so on. If you get a strong impression about the classroom try to briefly capture this and provide some evidence to support your impression.

From Considerations page: If seating arrangement changes during various segments of the observation, note these in the narrative section. If considerable various occur it may be necessary to draw a second sketch, otherwise, document the types of changes and where they occur in the descriptive section. If something not listed in the "condition of the room" is apparent, make note and discuss at first meeting with research team. Variables listed in "walls" section should be briefly described in the narrative section. The type of student assignments should be described in such a way to determine if only the best work is posted or other. Pictures, graphs, and charts should be described in the narrative section. Consideration should be given to whether these pictorials are teacher made, commercial, or student made; outdated, worn, new, etc. For "rules of behavior" the researcher should consider the tone of the rules, the number posted, if they teacher/student/school created (may have to asked teacher). In the "room contents[learning center]" section, researchers should briefly describe the number of centers as well as give brief descriptions of the centers' purpose (if apparent). If centers revolve around a theme or central idea, document this. If centers appear random/chaotic/exceptional, document this and provide evidence to support your opinion/viewpoint.

SUMMATIVE OBSERVATION ASSESSMENT INSTRUMENT

SUMMATIVE OBSERVATION ASSESSMENT

A. Basic Descriptive Information

1. Teacher observed: _____ 2. Observer's name:_____

3. Grade level(s): _____ 4. Subject observed: _____

Observation: 5. Date _____

6. Time: _____

7. Length: _____minutes

8. Class period/scheduled length of class (lesson): _____ minutes

Students:

Gender 9. _____ Male

10. _____ Female

Ethnicity 11. _____ African American

12. _____ Caucasian

13. _____ Hispanic

14. _____ Other

15. Copies of instructional materials collected? (attach to SOA) Yes No

16. Other adults (paraprofessional, assistant) in class? Yes No

B. Purpose and Focus of the Lesson

1. Purpose of the lesson (put in reflective survey also):

2.Focus of the lesson:
(This may be documented by segments. A teacher may cover (in 1 lesson) vocabulary, a reading strategy and comprehension. Other examples: language arts, phonemic awareness, writing, etc.)

a. Almost entirely working on 1 topic.

Specify: _____

b. Mostly working on 1 topic, but working some on other topics.

Specify: _____

c. About equally working on topic A and topic B.

Specify: _____

d. Other: _____

C. Lesson Ratings

Create phrases for referencing back to narrative of lesson. In narrative of lesson, mark phrases to support indicators. Example: Story mapping in small groups. (1.f)

(O =observed, N =not observed, DK = Don't Know, NA = Not applicable)

1. Planning

a. The lesson incorporated tasks and interactions consistent with the content observed.
(a) O N DK NA

b. The lesson reflected thorough planning and organization. (b) O N DK NA

c. The instructional strategies and activities observed in this lesson reflected the teacher's understanding of the students' levels of preparedness, prior knowledge and/or learning styles.
(c) O N DK NA

d. The resources used in the lesson supported the achievement of the instructional goals.
(d) O N DK NA

e. The strategies and activities observed reflected awareness of diversity and equity.
(e) O N DK NA

f. The lesson encouraged collaboration among the students. (f) O N DK NA

2. Implementation

a. The instructional strategies were appropriate for the content of the lesson observed. (What was being taught was possible through the instructional strategies observed)
(a) O N DK NA

b. The teacher demonstrated confidence in teaching the content of the lesson.
(b) O N DK NA

c. The instruction was paced to meet the developmental needs of the students.
(c) O N DK NA

d. The questioning strategies emphasized effective use of higher order questioning.
(d) O N DK NA

e. The questioning strategies emphasized appropriate use of "wait time."
(e) O N DK NA

f. Evidence of metacognition was present. (f) O N DK NA

3. Content

a. Content of the lesson was appropriate for all students in this class. (a) O N DK NA

b. The teacher demonstrates an understanding of the concepts taught/presented in the lesson observed through dialogue with her students. (b) O N DK NA

c. Student engagement focused on academic ideas that supported the lesson's objectives.
(c) O N DK NA

d. Content specific vocabulary was used and/or encouraged throughout the lesson.
(d) O N DK NA

e. Instruction included connections to the real-world and other contexts when appropriate/possible. (e) O N DK NA

4. Classroom Culture

a. Teacher encouraged and valued the active participation of all students.
(a) O N DK NA

b. Student interactions reflected appropriate working relationships. (b) O N DK NA

c. Evidence of intellectual rigor was present (constructive criticism, the challenging of ideas, etc.). (c) O N DK NA

5. Assessment

a. Students assessed themselves. (a) O N DK NA

b. Peer assessment was evident. (b) O N DK NA

c. Teacher used traditional assessment. (e.g., paper/pencil test, Q & A) (c) O N DK NA

d. Teacher used non-traditional assessment. (e.g., project, portfolio) (d) O N DK NA

e. Teacher assessed by walking around/monitoring. (e) O N DK NA

6. Behavior management

a. Teacher addressed/managed behavior according to behavioral expectations of the class.
(a) O N DK NA

b. Teacher addressed inappropriate behavior that was disruptive to the learning environment.
(b) O N DK NA

c. Behavioral expectations were posted/visible in the classroom. (c) O N DK NA

d. Time spent devoted to instructional activities _____ minutes

e. Time spent devoted to managing student behavior/dealing with discipline
_____ minutes

7. Lesson segments

Time	Segment	Teacher	Student	Grouping	Other
	1				
	2				
	3				
	4				
	5				
	6				
	7				
	8				
	9				

Teacher behaviors:

a	answering questions
b	asking questions
c	circulating
d	distributing materials
e	giving closure
f	guided reading
g	introducing
h	making assignments
I	presenting information
j	reviewing
k	summarizing
l	
m	
n	
o	

Student behaviors:

a	answering questions
b	asking questions
c	doing worksheets
d	listening attentively
e	not listening attentively/off task
f	paying attention to visuals
g	reading aloud
h	reading silently
I	working at the board
j	working in his seat
k	writing down assignments
l	writing/taking notes
m	
n	
o	

Groupings:

a	individual
b	paired
c	whole class
d	small group
e	large group
f	
g	
h	
I	

GENERAL REFLECTIVE SURVEY

GENERAL REFLECTIVE SURVEY

(PLEASE COMPLETE WITHIN 48 HOURS OF OBSERVATION)

Please reflect on the lesson observed by the member of the PEF research team. The following survey should be answered, addressing the concerns as they relate to (a) the actual lesson observed or (b) the subject area observed.

*Consider your thoughts, philosophies and teaching qualities, read through each question carefully and select the best answer. It is very important that **all** statements are answered. For any questions needing additional space, please feel free to attach a sheet with comments, making sure that comments are properly numbered.*

Name: _____ Subject Observed: _____

Grade(s): _____ School: _____

CLARITY: *(additional observer question)*

A. THE LESSON OBSERVED

Section A is specific to the **lesson observed** by the member of the PEF research team. Please respond to the questions/statements in respect to the experience **during** the lesson/activity.

1. a. What was the purpose of today's lesson?

   b. What concept(s) or skills did you want kids to learn in this lesson?

2. a. How successful was the lesson? (Circle the corresponding number with 0 being "not successful" to 7 being "very successful.")

Not successful Very successful

0 1 2 3 4 5 6 7

b. What worked?

c. What might you do differently next time?

3. Did you ***deliver/teach*** the lesson essentially as it was organized (without modifications)? If no, briefly describe the modifications you made and your reasons for making them.

☐ Yes

☐ No (briefly describe)

4. What particular instructional challenges did the students in the class offer?

☐ None

Briefly describe:

5. What resources did you use to plan this lesson?

☐ None

Briefly describe:

6. Did you choose the resources that you used yourself?

☐ No

☐ Yes

Describe and list the instructional materials used in the lesson.

7. Please help me understand where this lesson fits in the sequence of the unit you are presently teaching.
   a. What have the students experienced prior to today's lesson?

   b. Is today's lesson/activity part of a unit? If yes, where is the lesson/ activity situated in the development of that unit [e.g. day 1 (introduction) of 5 days needed to complete the unit]?

   c. What is the next step for this class in this unit?

B. COURSE SPECIFIC

Section B is specific to the subject/course observed by the member of the PEF research team. Please respond accordingly.

1. How are the students in the class observed assigned to you? (Mark an X on only one statement.)

   - ☐ All of the students in the class observed come from my self-contained classroom.
   - ☐ The students in the class observed come from two or more classrooms/homerooms of the same grade in this school.
   - ☐ The students in the class observed come from two or more classrooms/homerooms of different grade levels in this school.
   - ☐ Other (briefly explain): _____

2. a. How many times does the selected instructional period take place in a day?

   _____ per day

   b. How many times does the selected instructional period take place in a week?

   _____ per week

3. How many **minutes** long is this instructional period each day? (If instructional period does not meet daily, please explain.) _____ each day

4. How many students do you teach in your selected instructional period?

   _____ students

5. What is the ***main*** grade level of the students in this instructional period?

   _____ grade

6. How would you describe the ability level of the students in this instructional period?

(Mark only one item.)

- ☐ High ability
- ☐ Average ability
- ☐ Low ability
- ☐ Mixed abilities

C. INSTRUCTION

*Section C is specific to your **instructional practices** for the class observed. (Refer to the addendum to the reflective survey if you are unsure of the specific subject of your selected instructional period.)*

1. Approximately how often do you use each of the following teaching methods in the class observed? (Circle one number on each line.)

	0 = Never	**1 = 1-2 times a month**	**2 = 1-2 times a week**	**3 = Almost everyday**	**4 = Every class**
a. Lecture or talk to the whole class	0	1	2	3	4
b. Teacher-led whole class discussions	0	1	2	3	4
c. Students responding orally to questions on subject matter covered in class or homework	0	1	2	3	4
d. Student-led whole-group discussions or presentations	0	1	2	3	4
e. Students working together in cooperative groups	0	1	2	3	4
f. Reviewing homework or other assignments	0	1	2	3	4

2. In the class observed, how much emphasis do you give to each of the following goals or objectives? (Circle one number on each line.)

Emphasis on…	0 = None	1 = Minor	2 = Moderate	3 = Major
a. integrating the course curriculum with other subjects or fields of study	0	1	2	3
b. teaching facts, rules or vocabulary	0	1	2	3
c. showing the importance of the subject in everyday life	0	1	2	3
d. increasing students' interest in the subject and in pursuing further study	0	1	2	3
e. encouraging students to explore alternative explanations or methods for solving problems	0	1	2	3
f. preparing students for taking standardized tests in the subject	0	1	2	3
g. fully covering the course curriculum as prescribed by the school/district/state	0	1	2	3
h. in-depth study of selected topics or issues, as opposed to exposure to a broad range of topics	0	1	2	3

3. When teaching the observed class, how often do you use the following approaches to group students for instruction? (Circle one number on each line.)

	0 = Rarely or never	1 = A few times a month	2 = A few times a week	3 = Every day
a. Whole class grouping (i.e., all students are taught the same thing at the same time)	0	1	2	3
b. Ability or achievement grouping (e.g., the most proficient readers are in one group, the next most proficient are in a second group and the rest are in a third group)	0	1	2	3

	0 = Rarely or never	1 = A few times a month	2 = A few times a week	3 = Every day
c. Mixed ability grouping (e.g., students are grouped according to interest/genre, cooperative-learning groups)	0	1	2	3
d. Individualized instruction (e.g., students work individually on learning assignments specifically tailored to their achievement or interest)	0	1	2	3

4. Please indicate how confident you feel about the following aspects of your teaching of the subject observed by the PEF researcher. (Circle one number on each line.)

	0 = Not at all	1 = Slightly confident	2 = Moderately confident	3 = Very confident
a. Your knowledge about the application of the subject to everyday life	0	1	2	3
b. Your ability to advise students about opportunities to receive further training/experience in the subject area	0	1	2	3
c. Your ability to use inquiry-based instructional practices	0	1	2	3
d. Your ability to determine the depth, breadth and pace of coverage of material in your teaching	0	1	2	3
e. Your ability to develop appropriate and authentic assessment tools	0	1	2	3
f. Your ability to supervise research projects of your students	0	1	2	3
g. Your ability to mentor beginning teachers	0	1	2	3
h. Your ability to make presentations at teacher inservices or professional meetings	0	1	2	3
i. Your ability to incorporate technology (computers, the Internet, etc.) into your teaching	0	1	2	3

5. During a typical week, approximately how much time do you spend **outside of regular school hours** on planning and preparing for teaching this course? (Round the amount of **time** to the nearest ½ hour. Example: 2 hours and 20 minutes would be rounded to 2.5 **hours**.)

_____ hours per week

D. ASSESSMENT

Section D addresses assessment issues specific to the class observed. Please respond accordingly.

1. To what extent do you use the following types of assessment to determine student progress and achievement in this course? (Circle one number on each line.)

Type of assessment used:	0 = Not at all	1 = Slight extent	2 = Moderate extent	3 = Great extent
a. Pre-tests before beginning a new unit	0	1	2	3
b. Short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank)	0	1	2	3
c. Test requiring open-ended response (e.g., descriptions, justifications, explanations)	0	1	2	3
d. Student portfolios	0	1	2	3
e. Class participation/group discussion	0	1	2	3
f. Student presentations/projects	0	1	2	3
g. Hands-on performance measurements	0	1	2	3
h. Written explanations of thought processes (e.g., journals, essays)	0	1	2	3

2. Do you use the tests that the publishers included with the textbook/workbook? (Mark an X on one line only.)

- ☐ Rarely or never
- ☐ Sometimes
- ☐ Frequently

E. OTHER INFORMATION

*Section E **contains information on various topics**. Some are specific to the class observed others are specific to you. Please read each question carefully.*

1. Which subject/area is your favorite to teach? (Mark an X on only one subject/area.)

- ☐ None
- ☐ English
- ☐ Language arts
- ☐ Reading
- ☐ Mathematics
- ☐ Natural science (biology, chemistry, physics)
- ☐ Social science (psychology, sociology)
- ☐ History
- ☐ Fine arts (arts, music, drama)
- ☐ Physical education
- ☐ Other: specify _____

2. Were you *officially* mentored as a new teacher? (*Officially* is defined as a mentor designated by the district.)

- ☐ No
- ☐ Yes

If yes, this person was: (Please check all that apply.)

- ☐ in your subject area.
- ☐ at your grade level.
- ☐ on your hall.
- ☐ in your school.
- ☐ other (please specify) _____

3. Were you *unofficially* mentored as a new teacher? (*Unofficially* chosen by you or someone who chose you to mentor.)

- ☐ No
- ☐ Yes

If yes, this person was: (Please check all that apply.)

- ☐ in your subject area.
- ☐ at your grade level.
- ☐ on your hall.
- ☐ in your school.
- ☐ other (please specify) _____

4. Approximately what percentage of the class textbook/workbook do you typically cover in this course?

_____% of the textbook is typically covered

- ☐ There is no textbook/workbook for this course

5. Classroom computers
   a. How many computers are located in the classroom where you teach? _____
   b. How many *working* computers are located in the classroom where you teach? _____
      1. How many have Internet access? _____
      2. How many computers with Internet access are available to the students in the classroom where you teach? _____

6. Classroom printers
   a. How many printers are located in the classroom where you teach? _____
   b. How many *working* printers are located in the classroom where you teach?_____

7. Other classroom technology

What other technology is housed in your classroom for use by you and the students? Please list and explain each item.

8. Certifications

Teaching certification/licensure: (Mark an X on all that apply.)

- ☐ Elementary – general
- ☐ Elementary – specialist, Area(s): _____
- ☐ Secondary – middle school/junior high (indicate area(s) of certification below)
  - o Math
  - o English
  - o Science
  - o Social Studies
  - o Other (please specify): _____
- ☐ Secondary – high school (indicate area(s) of certification below)
  - o Math
  - o English
  - o Science
  - o Social Studies
  - o Other (please specify): _____
- ☐ Other certifications: Please be specific in describing other type(s) of certification/licensure you hold.

REFERENCES

Adams, M. (1990). Beginning to read: Thinking and learning about print. In T. L. Good and J. E. Brophy, *Looking in classrooms,* 9th edition (p. 411). Boston: Allyn and Bacon.

Airasian, P. W. (1997). Oregon teacher work sample methodology: Potential and problems. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 46-52). Thousand Oaks, CA: Corwin Press.

Baker, A. & Xu, D. (1995). *The measure of education: A review of the Tennessee value-added assessment system.* Nashville: Tennessee Department of Educational Accountability.

Baltimore County Public Schools Survey. No author: Division of Research, Evaluation, Assessment and Accountability. *Factors impacting staff employment decisions: survey results of current and former teachers, administrators, and additional staff.* A research study conducted for the Board of School Commissioners. (found electronically)

Ball, D. L. (1997). What do students know? Facing challenges of distance, context, and desire in trying to hear children. In B. Biddle, T. Good, & I. Goodson (Eds), *International handbook of teachers and teaching* (pp. 769-818). The Netherlands: Kluwer Academic Press.

Ball, D. L., Camburn, E., Correnti, R., Phelps, G., & Wallace, R. (1999). *New tools for research on instruction and instructional policy: A web-based teacher log.* A CTP working paper. University of Washington: Center for the Study of Teaching and Policy.

Bamburg, J. (1994). *Raising expectations to improve student learning.* In G. D. Borich, *Observation skills for effective teaching* (p. 66). Upper Saddle, NJ: Merrill-Prentice Hall.

Barr, A. S. (1929). *Characteristic differences in the teaching performance of good and poor teachers of the social studies*. Bloomington, Illinois: Public School Publishing Company.

Beecher, D. E. (1949). *The evaluation of teaching, backgrounds and concepts*. Syracuse NY: Syracuse University Press.

Bellon, J. J., Bellon, E. C. & Blank, M. A. (1992). *Teaching from a research knowledge base: A development and renewal process*. New York: Macmillan Publishing Company.

Biddle, B. J. & Ellena, W. J. (1964). *Contemporary research on teacher effectiveness.* New York: Holt, Rinehart and Winston.

Bock, R., Wolfe, F. & Fisher, T. (1996). A review of analysis of the Tennessee value-added assessment system. Nashville: Tennessee Department of Educational Accountability.

Boehm, A. E. & Weinberg, R. A. (1997). *The classroom observer: A guide for developing observational skills.* New York: Teachers College Press.

Borich, G. D. (2004). *Effective teaching methods.* New Jersey: Pearson.

Borich, G. D. (1996). Effective teaching. In P. R. Burden & D. M. Byrd, *Models for effective teaching* (p. 93). Boston: Allyn & Bacon.

Borich, G. D. & Madden, S. K. (1977). *Evaluating classroom instruction: A sourcebook of instruments.* Reading, MA: Addison-Wesley Publishing Company.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Brim, O. G. (1958). *Sociology and the field of education*. New York: Russell Sage Foundation.

Brophy, J. (1992). Probing the subtleties of subject-matter teaching. *Educational Leadership, 49*(7), 4-8.

Brophy, J. & Evertson, C. (1976). *Learning from teaching: A developmental perspective.* Boston: Allyn & Bacon.

Brophy, J. E. & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences.* New York: Holt, Rinehart & Winston.

Brown, G. & Wragg, E. (1993). Questioning strategies. In G. D. Borich, *Effective teaching methods* (p. 259). New Jersey: Pearson.

Brualdi, A. C. (1998). *Classroom questions* (Report No. EDO-TM-98-02). Washington, DC: Catholic University of America, Department of Educations, ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document No. ED ED422407)

Bryan, R. C. (1937). *Pupil rating of secondary teachers*. New York: Teachers College, Columbia University.

Burden, P. R. & Byrd, D. M. (2003). *Models for effective teaching.* Boston: Allyn & Bacon.

Burke, D. L. (1997). Looping: Adding time, strengthening relationships. In G. D. Borich *Observation skills for effective teaching* (p.95). Upper Saddle River, NJ: Merrill-Prentice, Inc.

Cabello, B. & Terrell, R. (1994). Making students feel like family: How teachers create warm and caring classroom climates. *Journal of Classroom Interaction, 29*, 17-23.

Campbell, R. J., Kyriakides, L., Muijs, R. D. & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education, 29*(3), 347-362.

Cantrell, S. C. (1998/1999). Effective teaching and literacy learning: A look in side primary classrooms. *The Reading Teacher, 52*(4), 370-378.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*(8), 723-733.

Cartwright, C. A. & Cartwright, G. P. (1974). *Developing observational skills.* New York: McGraw-Hill.

Center for the Study of Teaching and Policy (CTP). *CTP Teacher Survey*. Retrieved August 3, 2000 from: http://www.stanford.edu/group/CRC/survey_instruments.htm

Ceperley, P. E. & Reel, K. (1997). The impetus for the Tennessee value-added assessment system. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 133-136). Thousand Oaks, CA: Corwin Press.

Charters, W. W. & Waples, D. (1929). *The Commonwealth teacher-training study*. Chicago, Illinois: The University of Chicago Press.

Chuska, K. (1995). Improving classroom questions: A teacher's guide to increasing student motivation, participation and higher level thinking. In G. D. Borich, *Effective teaching methods* (p. 13). New Jersey: Pearson.

Collins, A. (1990). Transforming the assessment of teachers: Notes on a theory of assessment for the 21st century. *National Catholic Education Association*. Boston, MA: National Catholic Education Association.

Cohen, L. (1976). *Educational research in classrooms and schools: A manual of materials and methods.* London: Harper and Row.

Coker, H. & Coker, J. (1982). *Classroom observations keyed for effectiveness researcher (COKER).* Atlanta: Georgia State University.

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

Collins, A. (1990). Transforming the assessment of teachers: Notes on a theory of assessment for the 21st century. *National Catholic Education Association*. Boston, MA.

Cotton, K. (1996). Expectations and student outcomes. In G. D. Borich, *Observation skills for effective teaching* (p. 94). Upper Saddle River, NJ: Merrill-Prentice, Inc.

Cowan, T. (2003). *Value-added agricultural enterprises in rural developmental strategies.* New York, NY: Novinka Books.

Creemers, B. P. M. & Reezigt, G. J. (1997). School effectiveness and school improvement: Sustaining links. *School effectiveness and school improvement, 8*(4), 463-470.

Cruickshank, D. R. (1990). *Research that informs teachers and teacher educators.* Bloomington, IN: Phi Delta Kappa.

Cruickshank, D. & Metcalf, K. (1994). Explaining. In G. D. Borich, *Observation skills for effective teaching* (p. 12). Upper Saddle River, NJ: Merrill-Prentice, Inc.

Curwin, R. L. & Mendler, A. N. (1999). In G. D. Borich, *Observation skills for effective teaching* (p. 94). Upper Saddle River, NJ: Merrill-Prentice, Inc.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. & McGreal, T. (2000). *Teacher Evaluation to Enhance Professional Practice*. Princeton, NJ: Educational Testing Service.

Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *Phi Delta Kappan, 73*(3), 220-225.

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

Darling-Hammond, L. (1998). Teachers and teaching: Testing policy hypotheses from a national commission report. *Educational Researcher, 27*(1), 5-15.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 1-40.

Darling-Hammond, L. & Wise, A. (1985). Beyond standardization, state standards, and school improvement. *Elementary School Journal, 85*(3), 315-336.

Darlington, R. B. (1997). The Tennessee value-added assessment system: A challenge to familiar assessment methods. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 163-168). Thousand Oaks, CA: Corwin Press.

Denny, D. A., Rusch, R. & Ives, S. (1969). In A. Simon & E. G. Boyer (Eds.), Mirrors for Behavior III (pp. 219-222). Wyncote, PA: Communications Materials Center.

Doran, H. C. & Izumi, L. T. (2004). *Putting education to the test: A value-added model for California*. San Francisco, CA: Pacific Research Institute.

Dunkin, M. & Biddle, B. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.

Elliott, J. & Adelman, C. (1975). *Eliciting pupils' accounts in the classroom.* Norwich: Ford Teaching Project.

Elliott, J. (Ed.). (1993). *Reconstructing Teacher Education*. London: Falmer Press.

Elliott, J. (1988). *Teachers as researchers*. In J. P. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook* (pp. 78-81). Oxford, England: Pergamon Press.

Evertson, C., Emmer, E., Sanford, J., Clements, B., & Worsham, M. (1997). *Classroom management for elementary teachers.* Upper Saddle River, NJ: Prentice Hall.

Fallon, D. (2004). *Clarifying how we think about teaching and learning.* Speech presented at the 2004 National Value-Added Conference. Columbus, Ohio.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28*(2), 465-499.

Fitzpatrick, K. A. (1982). The effect of a secondary classroom management training program on teacher and student behavior. In D. Muijs and D. Reynolds, *Effective teaching: Evidence and practice* (pp. 94-95). Thousand Oaks, CA: Sage Publications.

Flanders, N. A. & Simon S. (1969). Teacher effectiveness. In R. L. Ebel (Ed.) *Encyclopedia of educational research* (4th ed.) (pp. 1423-1436). New York: MacMillan.

Flanders, N. A. (1970). *Analyzing teaching behavior.* Philippines: Addison-Wesley Publishing Company.

Fraser, B. & Walberg, H. (Eds.). (1991). Educational environments: Evaluation, antecedents and consequences. In G. D. Borich, *Observation skills for effective teaching* (p. 65). Upper Saddle River, NJ: Merrill-Prentice Hall, Inc.

Freiberg, H. J., Stein, T., & Huang, S. (1995). Effects of a classroom management intervention on student achievement in inner-city elementary schools. *Educational Research and Evaluation: An International Journal on Theory and Practice, 1*, 36-66.

Foster, P. (1996). *Observing Schools: A methodological guide.* London: Paul Chapman Publishing, Ltd.

Foster, P., Gomm, R., & Hammersley, M. (1996). Constructing educational inequality: An assessment of research on school process. London: The Falmer Press.

Gage, N. L. (1965). Desirable behaviors of teachers. *Urban Ed, 1*, 85-95.

Gage, N. L. (1972). *Teacher effectiveness and teacher education*. Palo Alto, CA: Pacific Books, Publishers.

Gage, N. L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.

Gagne, R. M. (1970). *Policy implications and future research: A response. Do teachers make a difference? A report on recent research on pupil achievement.* In Board of Education Personnel Development (Ed.) (pp. 169-173): Washington, DC: Office of Education.

Gall, J. & Gall, M. (1990). Outcomes of the discussion method. In W. W. Wilen (Ed.), *Teaching and learning through discussion: The theory and practice of the discussion method* (pp. 31-35). Springfield, IL: C. C. Thomas.

Galton, M. (1988). Structured observations. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (pp. 474-478). New York: Pergamon Press.

Germuth, A. (2003). Comparing results from value-added HLM and OLS models to assess teacher effectiveness. (Doctoral dissertation, University of North Carolina at Chapel Hill, 2003). *Dissertation Abstracts International, 64*(11), 4012.

Goldhaber, D. D. & Brewer, D. J. (2001). Evaluating the evidence of teacher certification: A rejoinder. *Educational Evaluation and Policy Analysis, 23*(1), 79-86.

Goldstein, H. (2001). *Using pupil performance data for judging schools and teachers: Scope and limitations*. London: University of London.

Good, T. (1979). Teacher effectiveness in the elementary school. *Journal of Teacher Education, 30*, 52-64.

Good, T. L. & Brophy, J. E. (1997). In J. Stronge, Q*ualities of Effective Teachers* (p. 40). Alexandria, VA: Association for Supervision and Curriculum Development.

Good, T., Brophy, J. & Biddle, B. (1975) *Teachers make a difference.* New York: Holt, Rinehart and Winston.

Good. T. & Grouws, D. (1979). Teaching effects: A process-product study in fourth-grade mathematics classrooms. *Journal of Teacher Education, 28,* 49-54.

Goodlad, J., Soder, R., & Sirodsnik, K. (Eds.). (1990). In T. L. Good & J. E. Brophy, *Looking in classrooms* (9th ed.) (p.25). Boston: Allyn & Bacon.

Gordon, H. R. & Yocke, R. (1999). Relationship between personality characteristics and observable teaching effectiveness of selected beginning career and technical education teachers. *Journal of Vocational and Technical Education, 16*(1), 47-66.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141-1177.

Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (1998). *Teachers, schools, and academic achievement* (pp.1-50). Cambridge, MA: National Bureau of Economic Research.

Haycock, K. (1998). Good teaching matters…A lot, Thinking K-16. Washington, D.C.: The Education Trust. (pp. 1-16). Retrieved August 3, 2000 from http://www.edtrust.org

Hook, C. (1981). *Studying classrooms.* Victoria: Deakin University Press.

Hunter, M. (1994). *Enhancing teaching.* Upper Saddle River, NJ: Merrill/Prentice Hall.

Hutt, S. J. & Hutt, C. (1970). *Direct observation and measurement of behavior.* Springfield, Illinois: Charles C. Thomas Publisher.

Jones, E. (1990). Interpersonal perception. In T. L. Good & J. E. Brophy, *Looking in classrooms* (9th ed.) (p. 70). Boston: Allyn & Bacon.

Jordan, H. R., Mendro, R. L. & Weersinghe, D. (1997). Cumulative and residual effects of teachers on future student academic achievement, *National Evaluation Institute*. Indianapolis, IN: CREATE.

Keeves, J. P. (Ed.). (1988). *Educational Research, methodology, and measurement: An international handbook.* New York: Pergamon Press.

Kennedy, M. M., Ball, D. L., & McDiarmid, G. W. (1993). *A study package for examining and tracking changes in teachers' knowledge.* East Lansing, MI: National Center for Research on Teacher Education.

Kingston, N. & Reidy, E. (1997). The Kentucky Instructional Results Information System meets the critics. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 228-240). Thousand Oaks, CA: Corwin Press.

Lambert, N. & McCombs, B. (1998). How students learn: Reforming schools through learner-centered education. In T. L. Good and J. E. Brophy, *Looking in classrooms* (9th ed.) (p. 411). Boston: Allyn and Bacon.

Lockwood, J. R. & McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics, 1*, 223-252.

McBer, H. (2000). A model of teacher effectiveness. Report to the Department for Education and Employment. Retrieved from http://www.teachernet.gov.uk/teachinginengland/detail.cfm?id=521

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

McLean, R., Sanders, W. & Stroup, W. (1991). A unified approach to mixed linear models. *The American Statistician, 45*(1), 54-64.

Medley, D. M. & Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), *Handbook of Research on Teaching* (pp. 247-328). Chicago: RAND McNally.

Medley, D. M. (1979). *The Development and Use of Observation Schedule and Record, Form 5V.* Charlottesville, VA: School of Education, University of Virginia.

Medley, D. M., Coker, H. & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach.* New York: Longman.

Medley, D. M. (1987). Evolution of research on teaching. In M. J. Dunkin (Ed.), *The International encyclopedia of teaching and teacher education* (pp. 105-113). New York: Pergamon.

Merriam, J. L. (1905). Normal school education and teaching efficiency. In D. E. Beecher, *The evaluation of teaching, backgrounds and concepts* (p. 5). Syracuse NY: Syracuse University Press.

Millman, J. (1997). *Grading teachers, grading schools*. Thousand Oaks, CA: Corwin Press.

Mitzel, H. E. (1960). Teacher effectiveness. In C. W. Harris (Ed.), *Encyclopedia of Educational Research* (3rd edition). New York: Macmillan.

Moore, I. (1995). In P. R. Burden & D. M. Byrd, *Models for effective teaching* (p. 93). Boston: Allyn & Bacon.

Mosteller, F., Light, R. J. & Sachs, J. A. (1996). Sustained inquiry in education: Lessons in skill grouping and class size. *Harvard Education Review, 66*, 797-842.

Muijs, D. & Reynolds, D. (1999). School effectiveness and teacher effectiveness: some preliminary findings from the evaluation of the Mathematics Enhancement Programme. *School Effectiveness and School Improvement, 11*(3).

Newby, T. (1991). In T. L. Good and J. E. Brophy, *Looking in classrooms* (9th ed.) (p. 238). Boston: Allyn and Bacon.

Page, D. P. (1885). *Theory and practice of teaching or the motives and methods of good school-keeping.* New York: A.S. Barnes & Company.

Parsad, B., Lewis, L. & Farris, E. (2000). Teacher preparation and professional development. *Education Statistics Quarterly, 3*(3).

Perkins, H. A. (1964). A procedure for assessing the classroom behaviors of students and teachers. *American Educational Research Journal, 1,* pp. 249-260.

Porter, A. & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the institute for research on teaching. *Educational Leadership, 45*(8), 74-85.

Poulson, B. W. (1965). Value added in manufacturing, mining, and agriculture in the American economy from 1809 to 1839. (Doctoral dissertation, Ohio State University, 1965). *Dissertation Abstracts International, 27* (01), 40.

Putnam, R. T. & Borko, H. (1997). Teacher learning: Implications of new views of cognition. In B. Biddle, T. Good, & I. Goodson (Eds.), *International handbook of teachers and teaching* (Vol. 2, pp. 1223-1296). The Netherlands: Kluwer Academic Publishers.

Raudenbush, S. & Byrk, A. (1986). A hierarchical model of studying school effects. *Sociology of Education, 59*(1), 1-17.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes.* Washington, D. C.: Economic Policy Institute.

Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R. W. M. Travers (Ed.), *Second Handbook of Research on Teaching* (pp.126-136). Chicago: Rand McNally.

Rosenshine, B. & Meister, C. (1992). The use of scaffolds for teaching higher-level

cognitive strategies. *Educational Leadership, 49*(7), 26-33.

Ryans, D. G. (1960). *Characteristics of teachers, their description, comparison, and appraisal: A research study*. Washington, D.C.: American Council on Education.

Ryans, D. G. (1964). Research of teacher behavior in the context of teacher characteristics study. In B. J. Biddle and W. J. Ellena (Eds.), *Contemporary research on teacher effectiveness* (pp. 67-101). New York: Holt, Rinehart and Winston.

SAS Institute Incorporated. Dr. William Sanders biography. Retrieved February 12, 2008, from http://www.sas.com/govedu/edu/bio_sanders.html

Sanders, W. L & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed methodology in educational assessment. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system: A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 137-162). Thousand Oaks, CA: Corwin Press.

Sanders, W. L. & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-255.

Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee at Knoxville.

Schalock, H. D., Schalock, M. & Girod, G. (1997). Teacher work sample methodology as used at Western Oregon State College. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 15-45). Thousand Oaks, CA: Corwin Press.

Schmidt, W. H., McKnight, C. C. & Raizen, S. A. (1997). *A splintered vision: An investigation of U. S. science and mathematics education.* Boston: Kluwer Academic Press.

Shanoski, L. A. & Hranitz, J. R. (1992). Learning from America's best teachers: Building a foundation for accountability through excellence. Paper presented at the Annual Meeting of the Association of Teacher Educators (72nd, Orlando, FL, February 15-19, 1992).

Simon, A. & Boyer, E. G. (Eds.). (1974). *Mirrors for Behavior III: An anthology of observation instruments.* Wyncote, PA: Communications Materials Center.

Slavin, R. E. (1997). In P. R. Burden & D. M. Byrd, *Models for effective teaching* (p. 93). Boston: Allyn & Bacon.

Soar, R. (1971). *Follow through classroom process measurement.* Washington, DC: Florida Educational Research and Development Council, Gainesville, Gainesville: Office of Education (DHEW). ERIC Document ED 113 288

Soar, R. & Soar, R. (1982). Content effects in the learning process. In D. C. Smith (Ed.), *Essential knowledge for beginning educators* (pp.156-192). Washington, DC: American Association of Colleges of Teacher Education.

Solomon, L. C., White, T., Cohen, D. & Woo, D. (2007). *The effectiveness of the teacher advancement program.* Santa Monica, CA: National Institute for Excellence in Teaching.

Spaulding, R. (1967). Coping Analysis Schedule for Educational Settings (CASES). *An introduction to the use of the coping analysis schedule for educational settings and S-T-A-R-S.* Durham, NC: Educational Improvement Program, Duke University.

Stallings, J. A. (1977). *Learning to look: A handbook of classroom observation and teaching methods.* Belmont, CA: Wadsworth.

Stalling, J. A. & Mohlman, G. G. (1988). Classroom observation techniques. In J. P. Keeve (Ed.), *Educational research, methodology and measurement: A international handbook* (pp. 469-474). Oxford, England: Pergamon.

Stapleton, J., LeFloch, K. A. C., Bacevich, A. E. & Ketchie, B. (2004). Researching education as it happens: Using classroom observations to generate quantifiable data. A paper presented by American Institutes for Research at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 2004.

Stephens, J. M. (1967). *The process of schooling*. New York: Holt, Rinehart & Winston.

Stevens, J., Estrada, S. & Parks, J. (2000). *Measurement issues in the design of state accountability systems.* Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000). ED 445 026

Stone, J. E. (1999). *What is value-added assessment and why do we need it?* Policy Brief, 1999, No. 99-11. Alexandria, VA: Foundation Endowment. Retrieved November 14, 2004 from http://www.heartland.org/pdf/21824j.pdf

Stronge, J. H & Tucker, P. D. (2000). *Teacher evaluation and student achievement*. Washington, D. C: National Education Association.

Stufflebeam, D. L. (1997a). Oregon teacher work sample methodology: Educational policy review. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 53-61). Thousand Oaks, CA: Corwin Press.

Stufflebeam, D. L. (1997b). Overview and assessment of the Kentucky instructional results information system. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 219-227). Thousand Oaks, CA: Corwin Press.

Science Work Experience Programs for Teachers (SWEPT) pre-program survey. Retrieved August 3, 2000 from http://www.ncrrsepa.org/Eval/HSResearch/PreTeacherSurvey.pdf

Sykes, G. (1997). On trial: The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 110-119). Thousand Oaks, CA: Corwin Press.

Taylor, B. M., Pearson, P. D., Clark, K. F., & Walpole, S. (1999). Center for the improvement of early reading achievement: Effective schools/accomplished teachers. *The Reading Teacher, 53*(2), 156-159.

Teddlie, C. & Stringfield, S. (1993). *Schools make a difference: Lessons learned from a 10-year study of school effects.* New York: Teachers College Press.

Thum Y. M. & Bryk, A. S. (1997). Value-added productivity indicators: The Dallas system. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 100-109). Thousand Oaks, CA: Corwin Press.

Thum, Y. M. (2003). Measuring progress toward a goal: Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods and Research, 32*(2), 153-207.

Travers, R. M. W. (Ed.). (1973). *Second handbook of research on teaching: A project of the American Educational Research Association.* Chicago: Rand McNally.

Valli, L., Raths, J., & Rennert-Ariev, P. (2001). *A beginning teacher survey study: A theoretical perspective.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA, April 10-14, 2001.

Vogt, C. G. (1984). Developing a teacher evaluation system. *Spectrum, 2*(1), 41-46.

Walberg, H. J. & Paik, S. J. (1997). Assessment requires incentives to add value: A review of the Tennessee value-added assessment system. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 169-178). Thousand Oaks, CA: Corwin Press.

Wallen, N. E., Moohr, M., Hall, S. P. & Weisberg, K. (1969). Step Observation Schedule

(STEPOS). In A. Simon & E. G. Boyer (Eds.), Mirrors *for Behavior III: An anthology of observation instruments* (pp. 643-648). Wyncote, PA: Communications Materials Center.

Wallen, N. E. & Travers, R. M. W. (1963). Analysis and investigation of teaching methods. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 448-505). Chicago: Rand McNally.

Wang, M., Haertel, G. D., & Walberg, H. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 240-294.

Webster, W. J. & Mendro, R. L. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 81-99). Thousand Oaks, CA: Corwin Press.

Webster, W. J., Mendro, R. L., Orsak, T. & Weerasinghe, D. (1998). *An application of hierarchical linear modeling to the estimation of school and teacher effect.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 1998.

Weinstein, R. S. & McKown, C. (1998). In T. L. Good & J. E. Brophy, *Looking in classrooms* (9th ed.) (p. 25). Boston: Allyn & Bacon.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 Mathematics and Science education in the United States.* Report prepared with the support from the National Science Foundation under grant number REC-9910967. Chapel Hill, NC: Horizon Research, Inc.

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into the discussions of teacher quality.* A policy information report. Princeton, N.J.: The Milken Family Foundation and Educational Testing Service.

Wilen, W. (1991). In Burden, P. R. & Byrd, D. M., *Models for effective teaching* (p. 94). Boston: Allyn & Bacon.

Withall, J. (1949). The development of a technique for the measurement of social-emotional climate in classrooms. *Journal of Experimental Education, 17,* 347-361.

Withall, J. (1960). Research in systematic observation in the classroom and its relevance to teachers. *Journal of Teacher Education, (23)*3, 330-332.

Wilson, B. & Corbett, D. (1990). Statewide testing and local improvement: An oxymoron? In J. Murphy (Ed.), *The educational reform movement of the 1980s* (pp. 243-263). Berkeley, CA: McCutchan.

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