CONLEY, APRIL L., Ed.D. Teacher Perceptions of the North Carolina Educator Evaluation System and the Use of Value-Added Data as a Measure of Accountability. (2015)
Directed by Dr. Carl Lashley. 195 pp.

This qualitative research study is a phenomenological exploration of teachers' understanding of and perceptions about the teacher evaluation process in North Carolina and the use of the North Carolina Educator Evaluation System (NCEES). Twenty-three teachers with varying years of experience and from six schools of varying demographics were interviewed about their knowledge and perceptions of each evaluation standard and were asked to provide examples of how they have used feedback from the evaluation process to inform their professional practices. Overall, participants had a limited understanding of the standards-based portion of the NCEES and of the value-added data component of the NCEES. Teachers received limited feedback from the evaluation process and were generally unable to provide examples of how they have been able to use feedback from the evaluation process to inform their professional practices. Teachers provided insight about the strengths and weaknesses of the NCEES and suggestions for improvement in the evaluation process and in the NCEES. As a result of the findings, implications for future teacher evaluation in North Carolina are discussed.

TEACHER PERCEPTIONS OF THE NORTH CAROLINA EDUCATOR EVALUATION SYSTEM AND THE USE OF VALUE-ADDED DATA AS A MEASURE OF ACCOUNTABILITY

by

April L. Conley

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Greensboro
2015

Approved by

Carl Lashley
Committee Chair

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

I dedicate this endeavor to a dear friend. Dennis, your influence has shaped who I am today and, therefore, the footprints I leave are partially left by you as well. I am a better person as a result of having known you. We started this journey together, and you are still with me.

APPROVAL PAGE

This dissertation, written by April L. Conley, has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair Carl Lashley

Committee Members Craig Peck

Rick Reitzug

Leila Villaverde

March 5, 2015
Date of Acceptance by Committee

March 5, 2015
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to thank my mom and dad for providing me with a great home and childhood and for instilling in me a love for learning and an appreciation for all people. Both of them were public educators and it is impossible to say how many lives they touched over the course of their careers. I would also like to thank my family and friends who have loved and supported me and given me the encouragement to accomplish my goals.

I would also like to thank all of the great teachers in the public schools who taught me over the years. I continue to learn from great teachers. There are many out there and I thank all of them. I also thank Dr. Lashley and my committee members who most certainly belong among the "greats."

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Emergence of Accountability in Public Education

Throughout the history of public education, many reforms have reshaped public schools in the United States. However, since the 1980's and the publication of *A Nation at Risk* by the National Commission on Excellence in Education, politicians and the general public have sought transformational reform in America's public schools in response to disappointing educational outcomes for many of our youth (National Commission on Excellence in Education, 1983; Spring, 2011). Numerous reform ideas have surfaced since this time in hopes that the quality of the teaching force will improve and, as a result, significant reform agendas have passed through federal legislature. Two of the most recent and noteworthy pieces of legislation were the *No Child Left Behind Act of 2001* and the *American Recovery and Reinvestment Act of 2009*, commonly known as *Race to the Top*. States such as Colorado, Ohio, and California have also passed legislation that aims to reform schools and increase teacher effectiveness. Specifically, these states have changed the way teachers are evaluated, using value-added student growth data to measure teacher effectiveness (Pascopella, 2006). Some states are also utilizing a pay-for-performance system that not only ties teacher evaluation to student outcomes but ties teacher salaries to student outcomes as well (Goldstein, 2011; Toch, 2011).

Professional development reform, reform in teacher preparation programs, reform in mentoring programs, and many reforms in teacher practices, methods, and curriculum have also been tried across the country. Currently, many states are adopting the Common Core, a national curriculum designed to increase expectations and rigor, make curricula consistent from state to state, and make teaching and learning relevant to students in the twenty-first century (Stover, 2011). Other reform ideas in education include reduced class sizes, establishing school voucher programs, and opening charter schools (Cavanagh, 2011; Schachter, 2010; Stover, 2011). Some districts have utilized bonus pay for teachers who agree to work in hard-to-staff schools. Others have adopted alternative licensure pathways to recruit untraditional teaching professionals with valuable life and work experiences (Jacob, 2007). Despite these and other initiatives, student outcomes have continued to disappoint many citizens, politicians, and business leaders over the past several decades. As a result, the concept of accountability in public education has emerged.

The implementation of heightened accountability in our nation's schools is unfolding in North Carolina as it is across the nation, though accountability measures and processes vary from state to state. The North Carolina public school system has implemented a series of reform efforts since the early 1980's to improve the quality of education (Coble, 2000). The most noteworthy reform efforts related to accountability for teachers include the adoption of the School Improvement and Accountability Act of 1989, which provided more flexibility in decision making at the school level in exchange for greater accountability. This accountability included local plans for school

improvement and a report card system for schools. Then, in 1992, the General Assembly approved the Performance-Based Accountability Program, which required every school to have an improvement plan and the establishment of student performance indicators. The ABC's of Public Education followed in 1995, which provided incentive funding for schools achieving high student growth in reading, writing, and math. Consequences for schools not achieving adequate student growth in these areas were also outlined in the ABC's, which included the development of school assistance teams. Though the teams were supposed to help build capacity among staff members in low-performing schools, the teams were also assigned evaluative duties and to report back to the State Board of Education on low-performing schools' progress (State Board of Education, 2001).

In 2010, many states across the nation competed for federal funding from the Race to the Top grant under President Obama's administration. North Carolina was one of the states to receive funds. North Carolina's grant application included accountability measures for teachers in the form of an additional standard being added to the North Carolina Educator Evaluation System (NCEES) beginning in the 2010-2011 school year based on Education Value-Added Assessment System (EVAAS) data. The application stated that the evaluation system helps to "differentiate teachers and principals by identifying those at the top end…as well as those at the lower end who are in need of remediation or possible dismissal" (Race to the Top Application, 2010, p. 132). The application went on to state, "evaluations of teacher and principal effectiveness must contain, as a major component, assessment of a teacher's or principal's effect on the academic growth of his or her students" (p. 133). Never before in North Carolina's

history of public education had teachers' or principals' accountability for student growth been so clearly stated or their evaluations so directly linked to student outcomes.

Teacher Quality

There is no question that some teachers achieve better results for students than other teachers. A study conducted by Sanders and Rivers (1996) found that students who had three consecutive teachers having high value-added data gained 50 more percentile points on an achievement test than did students who had three consecutive teachers who fell within the average range. Hanushek (1992) found that students assigned to the lowest performing teachers only make a half year's growth in reading, while students assigned to the highest performing teachers make one and a half years' growth in reading. A study conducted by Stronge, Ward, and Grant (2011) found similar results. The students in their study who were taught by teachers falling in the highest quartile of value-added data performed over 30 percentile points higher in both reading and math than students taught by teachers in the lowest quartile of value-added data.

Some researchers contend that the effects of teacher quality can be further highlighted by comparing school systems internationally. Klein (2011) highlights the fact that the highest performing school systems in the world recruit teachers from the top third of college graduates, whereas the lowest performing school systems recruit the majority of their teachers from the bottom third of graduates. In the U.S., approximately half of the U.S. teaching force comes from the bottom third of performers on the SAT, and many of these teachers end up teaching in predominantly low-income schools. The result is a significant achievement gap between economically disadvantaged students compared to

their peers (Kristof, 2011; Morgan, 2012). Leading countries such as Singapore, Japan, Finland, and Sweden ensure that teachers are highly trained and accessible to the least advantaged students (Darling-Hammond, 2008; Morgan, 2012). Each of these countries is outperforming the U.S. on international tests of student achievement. For example, on the Program of International Student Assessment (PISA) 2012 math assessment, 29 countries or systems scored better than the U.S. Twenty-two systems outperformed the U.S. in science on the PISA. Countries that have consistently scored better than the U.S. in reading literacy on the PISA include, but are not limited to, Australia, Canada, Finland, Japan, Korea, New Zealand, Norway, and Singapore. Nineteen systems outperformed the U.S. in reading on the 2012 PISA. Since 2009, over 20 systems improved in literacy. The U.S. was not one of them (Kelly, Xie, Nord, Jenkins, Chan, & Kastberg, 2013).

In summary, various measures demonstrate that students in the United States are not performing as well as many of their counterparts worldwide. Given the importance of teacher quality on student performance and that highly effective teachers produce significantly more gains in student growth than ineffective teachers, the conclusion that critics draw is that many of our students are being taught by ineffective teachers or, at least, by less effective teachers than is necessary to adequately prepare students for graduation and career readiness in the twenty-first century of our global marketplace (Morgan, 2012).

Teacher Evaluation

While an abundance of research on teacher quality and the effects of inequitable access to quality teaching exists in the professional literature, many researchers such as

Hanushek (2007) argue that common measures of teacher quality are "largely uncorrelated with true quality" (p. 574). Debate exists on whether teacher effectiveness should be measured by teachers' qualifications, their instructional practices and delivery, their effects on student achievement, or a combination of these (Lewis et al., 1999; Stronge et al., 2011). Researchers have consistently found that a teacher's qualifications or experience levels have no correlation with student outcomes in the area of achievement (Hanushek, 2007; Jacob, 2007; Stronge et al., 2011). However, when comparing teachers who consistently produce the greatest amount of achievement growth in students, teachers in the highest quartile have fewer class disruptions, they have higher expectations for student behavior, they establish more routines and procedures, they have more positive interactions with students, and they require students to take a greater degree of responsibility in the classroom (Stronge et al., 2011). As a result of these findings, Stronge et al. (2011) offer a conceptual framework for understanding teacher quality. They identify four areas in which to evaluate teachers. These areas include: instruction, assessment, the learning environment, and personal qualities.

In North Carolina, the purpose of the teacher evaluation process is to assess teacher performance in relation to the North Carolina Professional Teaching Standards and to create an individual plan for professional growth (NC State Board of Education, Policy ID Number TCP-C-004). The teaching standards consist of skills in the areas of leadership, establishing a learning environment for all students, knowing content, facilitating learning, and reflecting on one's practice. In 2011, a sixth standard was added based on contributions made to students' academic success. This final standard is based

on student growth data (North Carolina Department of Public Instruction [NCDPI], 2013).

According to Shakman, Riordan, Sanchez, DeMeo Cook, Fournier, and Brett (2012), five states in the U.S. meet all criteria for having a required evaluation system for all general educators, having an evaluation system that is operational statewide, including multiple rating categories, and uses multiple measures of teacher effectiveness. North Carolina is one of those five states and only one of two whose evaluation system reflects all ten teaching standards of the Interstate Teacher Assessment and Support Consortium. These standards include: learner development, learner differences, learning environment, content knowledge, application of content, assessment, planning for instruction, instructional strategies, professional learning/ethical practices, and leadership/collaboration.

North Carolina's evaluation process was implemented statewide during the 2010-2011 school year. Standard One of the NCEES evaluates a teacher based upon his or her leadership skills in the classroom, the school, and the profession. The focus is also on teachers advocating for positive changes in policies and practices affecting students' learning. Ethical behavior is also included in Standard One.

Standard Two of the NCEES examines teachers' ability to provide a positive and respectful environment for students and their knowledge and integration of culturally relevant materials, contributions, points of view, and influences. Standard Two also examines a teacher's ability to maintain high expectations for students, appreciate their

differences, and adapt their practices for students with special needs. This standard also recognizes the importance of teachers collaborating with the families of their students.

Standard Three focuses on the content of instruction. Teachers are expected to align their instruction with the adopted curriculum. They are also expected to know the subject matter beyond the content they teach and to recognize the interconnectedness of their content area with other areas and disciplines. Finally, they are expected to make instruction relevant for students and to create the link between what they teach and twenty-first century skills.

Standard Four of the NCEES evaluates a teacher's knowledge of their students' development levels and learning needs and to differentiate their instruction accordingly. It also examines teachers' ability to use a variety of instructional methods, including technology, and to help students develop critical thinking skills and leadership qualities. Teachers' communication skills and their use of assessment methods are also included in this standard.

Standard Five of the NCEES examines teachers' ability to analyze student learning, to link their professional growth to their professional goals and to function effectively in a complex learning environment. In essence, this standard evaluates teachers' ability to think critically about their students' learning and their own practices and to adapt their practices as necessary based upon their reflections.

Finally, Standard Six of the NCEES assesses teachers' effect on student progress. This standard uses EVAAS data to examine students' measurable progress. Teachers are

identified as exceeding expected growth for his or her students, meeting expected growth, or not meeting expected growth based on EVAAS.

Value-Added Data

Value-added data (VAD) is a comparison between how much growth the student is expected to make and how much growth the student actually makes. The difference between expected and actual growth is considered to be the teacher's effect on the student. This effect is determined through a statistical procedure known as multiple regression (Winters, 2012). Multiple regression uses information such as previous test performance, socioeconomic status, English proficiency, class size, race, and gender to predict a student's performance and a teacher's effect is measured against this prediction, either exceeding the expectation or not (Barile, 2013; Duffrin, 2011; Green, Baker, & Oluwole, 2012; Winters, 2012). While the concept of measuring a student's actual growth against his or her predicted growth may seem like a straightforward concept, there is much debate over VAD and its use in teacher evaluations.

Proponents of VAD cite various reasons why it is a useful tool in the teacher evaluation process. They argue that the data provide evidence that students have learned the material taught to them and, thus, VAD is an objective measure of how effective the teacher has taught (Duffrin, 2011). Also, teachers' value-added data are much stronger predictors of how students will perform than other teacher factors currently emphasized, including years of experience and level of pay, or student factors that we emphasize as predictors, including socioeconomic status (Stronge et al., 2011). Because value-added measures specify the amount of growth a teacher's students have experienced, the

method is more fair than judging a teacher's performance based on the number of students identified as proficient, since not all students enter the classroom at an equal starting point (Good, Wiley, & Sabers, 2010). Supporters argue that VAD clearly separates the most effective teachers from the least effective ones. For example, Hanushek (2007) found that teachers with high value-added scores typically achieve 1.5 years of growth with their students, whereas teachers with negative value-added scores often achieve only one-half year of growth with their students. In addition, observable qualities identified in teacher quality research that tend to differentiate between effective and ineffective teachers correspond with value-added data. For example, teachers in the bottom quartile of value-added measures have disruptive behaviors occur in their classrooms three times as often as teachers performing in the top quartile (Stronge et al., 2011). These researchers also found that teachers in the top quartile had more routines and procedures, were more organized and prepared, communicated higher expectations, and had more positive interactions with students.

Those against the use of value-added data in teacher evaluations argue that the method has several flaws. First, there is more than one statistical model to determine a teacher's value-added data, which is one reason why its use is controversial; a teacher's value-added data can look different depending on which formula is used (Di Carlo, 2012; Green et al., 2012). Good et al. (2010) and Green et al. (2012) argue that value-added formulas are based on the assumption that students are randomly assigned to classes, but that is often not the case since students are often placed strategically in classes at many schools. Also, if too few student scores are used to calculate a teacher's value-added data,

then there is an increase in the possibility that the results are due to chance factors (Di Carlo, 2012; Duffrin, 2011). Green et al. (2012) argue that value-added measures are too statistically weak and have too much error due to extraneous factors, which is why multiple years of data are needed. They argue further that only a third of teachers in the highest or lowest quintiles actually stay in those quintiles from one year to the next. They also argue that even after three years of data, the error rate can still be as high as 25%. Goldstein (2011) cites factors such as parental involvement, tutoring, student motivational factors, and life events as factors that cannot be accounted for through the use of value-added data. Finally, those who are against the use of value-added data argue that most teachers do not teach subject areas or grades that have tests on which to base value-added data (Galley, 2011).

Rationale for the Study

The NCEES has been fully implemented for only four years. This instrument uses a combination of standards-based evaluation and value-added data to evaluate teachers. The use of value-added data to measure teacher effectiveness is currently one of the most controversial topics in educational policy discussions within the United States (Di Carlo, 2012). Teachers' understanding and experiences with the NCEES should have some impact on teacher practices in the classroom, but the impact is unknown. If teachers have a poor understanding of the five standards or of value-added data, this will affect how they respond to the evaluation process and to what extent they will use the standards-based feedback and the value-added data to reflect upon their professional practices. In addition, understanding how teachers perceive the evaluation process may provide insight

into what flaws, if any, exist with the instrument and how concerns may be addressed as we move forward with North Carolina's evaluation policies and practices and with the continued focus on accountability.

The purpose of this study is to explore teacher perceptions of performance accountability measures within the context of North Carolina's new teacher evaluation system. Specifically, this study seeks to examine how teachers' experiences with the NCEES affect their classroom practices. This study also examines the understanding and perceptions teachers have of the North Carolina Educators Evaluation System (NCEES), including the recent addition of value-added data and Standard Six of the NCEES. Finally, this study explores ideas teachers have as alternative ways to evaluate teacher performance. Therefore, the study answers five questions that focus on both teachers' understanding and their perceptions about the teacher evaluation process in North Carolina.

Overarching Research Question

1. What are teachers' experiences and perceptions about the evaluation process using the NCEES and how do these experiences and perceptions inform their practices?

Additional Research Questions

2. What do teachers understand about Standards One through Six on the NCEES?
3. What do teachers think about the fairness of standards-based portion of the NCEES?

4. What do teachers think about the fairness of value-added data as a component of the NCEES?
5. What recommendations do teachers offer about each standard, the NCEES, and methods of accountability for student performance?

The research questions were initially explored within the working contextual framework illustrated in Figure 1. This contextual framework is considered a working framework because it had the potential to evolve during the study. The framework adopts the contextual framework on teacher quality by Stronge et al. (2011) as it relates to the NCEES and the teacher evaluation process in North Carolina.

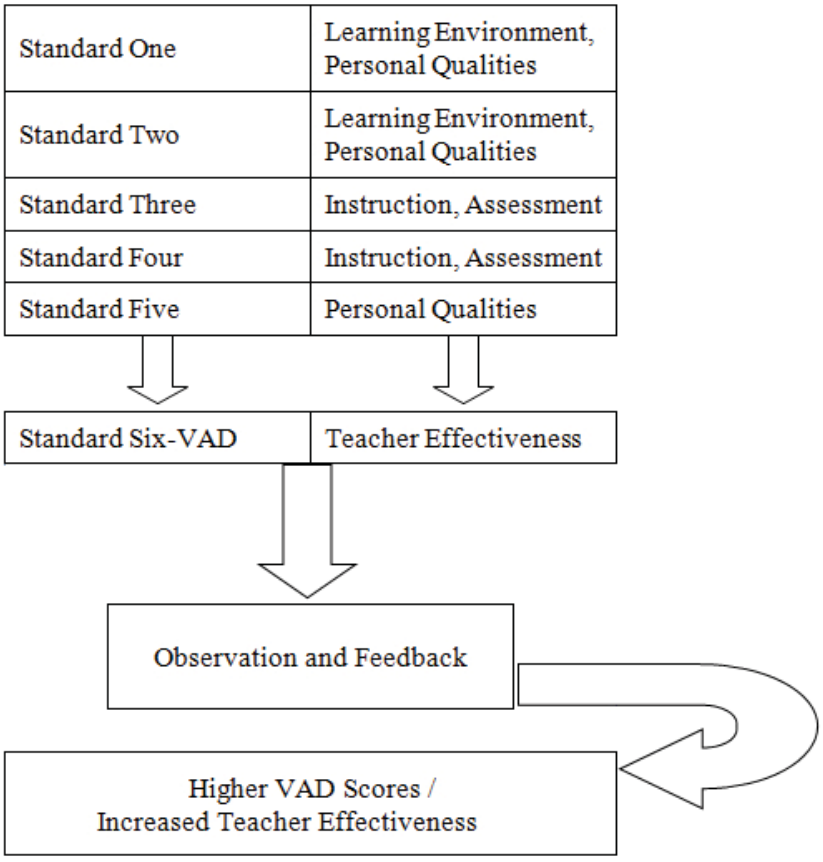

Figure 1. Working Contextual Framework.

Methodology

Research Tradition

This study is a qualitative research design. The purpose of the phenomenological approach is to explore teachers' understandings and perceptions of accountability and teacher evaluation and, specifically, of the NCEES through the use of semi-structured interviews with participants. The desired outcome is to make sense of teacher perspectives about teacher evaluation and how the process informs practice. Participants represent a variety of teacher demographics. I gathered relevant data through the semi-structured interviews and then examined participants' responses for patterns in their understanding and perceptions. I codified patterns and identified categories and emerging themes from the analysis of responses.

Key Concepts and Terms

Certain key concepts and terms need to be defined for the purpose of this study. Some of these key concepts and terms include the differentiation between types of schools and experience levels of teachers. Value-added data also needs to be clearly defined. Some of the key concepts and terms relative to the study are included here.

Experienced teacher—A teacher with five to fifteen years of teaching experience

High performing school—A school performing, at minimum, at the level of "distinction" and "high growth" on the most recent North Carolina school report card

Low performing school—A school performing as a "priority school" or "low performing" and "expected growth not achieved" on the most recent North Carolina school report card

Novice teacher—A teacher with five years or less of teaching experience

Standards-based evaluation—Evaluation of teacher performance based on classroom observations in which the evaluator compares a teacher's practices to well-defined professional standards

Value-added data—Data derived from statistical methods create predictive values about student growth, controlling for external factors other than the teacher's effect on growth. The predictive growth is then compared to actual growth and the difference is considered the teacher's effect.

Veteran teacher—A teacher who has more than fifteen years of teaching experience.

Setting

Several schools from a large school district in the Eastern United States are the settings of the study. Specifically, three elementary schools, two middle schools, and one school that serves kindergarten through eighth grade are used as settings for the study.

Participants

Participants include teachers from elementary and middle school levels. Specifically, teachers working in grades four through eight who receive value-added data reports were asked to participate, since these teachers have their own value-added data. The reason for including these teachers is based on the assumption that teachers with their own value-added data to populate Standard Six (as opposed to school-wide data) would have more input about how the data may affect their practices. Participants varied by educational background, levels of experience, performance ratings on the NCEES, and

school categorization. Twenty-three participants were included in the study in hopes of reaching saturation with the collected data.

Data Collection

Teachers in grades four through eight from three elementary schools, two middle schools, and one K-8 school were invited to participate in an interview of approximately 90 minutes in length, with the possibility of a follow-up interview. See Appendix A for the letter of invitation. See Appendix B for the teacher interview protocol. I utilized a guided, or semi-structured, interview approach, whereby I used the interview protocol as a guide, but I allowed myself to vary questions is response to the interview situation. I began each interview by developing rapport with the participant by means of thanking them for the volunteerism and reassuring them of confidentiality and anonymity. I was also cognizant of my nonverbal body language, including my body posture and facial gestures. I made a conscious effort to convey an informal and relaxed demeanor, with an interest in the participant's comments.

Data Analysis

I transcribed each interview from the audio recording and allowed each participant the opportunity to check the transcription for accuracy. I codified transcriptions to determine common themes for the purpose of a qualitative analysis of participant responses. I asked two knowledgeable colleagues to review my codifications and identified themes to confirm objectivity and accuracy. I used rich description to report emerging themes from the data.

Subjectivity

I am an administrator who has evaluated teachers using the NCEES and, prior to this study, I supported the use of value-added data in teacher (and all educator) evaluations, when used in conjunction with standards-based performance ratings. Therefore, my professional lens through which I initially analyzed the qualitative results about both standards-based and performance-based evaluation components of the NCEES were subjective based upon my professional experiences and current function within public schools. I attempted to address this subjectivity in two ways. First, interview questions were viewed by two professionals who have concerns with the use of value-added data in teacher evaluations and feedback was provided about the objectivity of items and the appropriate wording of items to ensure neutrality in the interview questions. Secondly, the peer reviewers checked interview transcripts and reviewed the codifying of responses and emerging themes from the data to ensure accuracy and objective reporting. However, as will be discussed in Chapter III, I experienced a significant change in perspective as a result of my findings from the study and my reflections upon them.

Trustworthiness

I utilized several strategies to ensure the credibility of the study and its findings. First, I collected sufficient data by interviewing 23 teachers from varying schools and grade levels. This provided a variety of perspectives and experiences. Also, I used the peer review method in which two knowledgeable colleagues analyzed my interpretations and checked for fairness and accuracy. I included rich description in my analysis of all data, clarifying my own researcher bias and clearly reflecting upon my own biases and

interpretations. I also used member checking, in which participants review their own interview transcription and are given the opportunity to view codifications to check for objectivity and accuracy.

All interview transcriptions were kept for auditing purposes. I audiotaped interviews. I recorded all dates and timelines of written materials, interviews, or other contacts with participants of the study in a journal. Copies of everything were kept secure in a locked location and all electronic files were password protected. I kept an extra copy of everything in a different location, also secured by physical lock or electronic password.

Benefits and Risks

With a better understanding of the perceptions of teachers on the issue of teacher accountability and evaluation, decision-makers will be better able to address specific concerns of teachers. The benefit of understanding perceptions of teachers on the issue of teacher accountability is that decisions will be made with the consideration and representation of individuals who are elemental in student achievement. Ultimately, educational decisions must be made on what the collective whole deems most beneficial to students. This study helps to uncover how the evaluation process ultimately affects teacher practices for the benefit of students. This study also added to the body of literature related to the debate surrounding the use of value-added data in teacher evaluations and provides insight into teacher perspectives on this important issue. Additionally, this study provides insight into the new NCEES and, specifically, how North Carolina policy surrounding teacher evaluation is translated into practice.

A potential risk of this study was the possible discomfort some teachers may have discussing their honest opinions and beliefs about accountability and the teacher evaluation process. Because teachers have increasingly become under public scrutiny, they were at risk to feel reluctance toward admitting to any negative perceptions they may have had of the NCEES or to any opposition they may have had related to the use of value-added data or to some other component of the NCEES for fear of being judged as uncommitted to student outcomes or for placing their own needs above those of students. These risks were addressed through open communication about the use of the results and guaranteed anonymity of participants.

Limitations

Participants were only chosen from elementary and middle schools within one district in North Carolina. Therefore, results of the study do not necessarily provide insight into the perspectives of teachers from other grade levels or from other districts or states.

Summary

This chapter provided a brief overview of relevant events in the history of public education in the United States that have collectively resulted in the emergence of accountability in public education. This chapter also discussed some important findings in teacher quality research in terms of differences in student outcomes for children taught by highly effective teachers compared to children taught by less effective teachers. Identified outcomes of variation in teacher quality include a resulting achievement gap between economically disadvantaged students compared to non-disadvantaged peers as

well as an overall mediocre performance by the U.S. on international testing, such as on the PISA.

Teacher evaluation should measure qualities that are found in teachers producing the most student achievement growth. Stronge et al. (2011) provide a conceptual framework for understanding these qualities. The framework, introduced in this proposal, includes four dimensions: instructional delivery, student assessment, learning environment, and personal teacher qualities. The more effective a teacher is in these domains should ultimately lead to more student growth and more positive value-added data. Researchers such as Hanushek (2007) and Stronge et al. (2011) believe this to be the case. However, those who argue against using value-added data in teacher evaluations contend that it is much more complex than this. This proposal briefly discussed both sides of this debate.

Since many states, including North Carolina, are adding value-added data to teacher evaluations, this chapter identified the significance of understanding how well teachers understand value-added data since teachers' understanding of it will impact the extent to which they will use the information to reflect upon their own practices. Also, understanding how teachers perceive the evaluation process and the inclusion of value-added data in the evaluation instrument may provide insight into how concerns may be addressed as the NC evaluation process continues to be refined. Thus, the purpose of this study is to explore teacher perceptions of performance accountability measures within the context of North Carolina's new teacher evaluation system. This chapter clarified that this study seeks to examine the understanding and perceptions teachers have of the North

Carolina Educators Evaluation System (NCEES), including the understanding and perceptions of the use of value-added data for Standard Six of the NCEES. This study also seeks to explore how feedback from each standard on the NCEES informs teacher practices and what suggestions or alternatives teachers offer as more meaningful ways to evaluate teacher effectiveness.

Finally, this chapter outlined the methodology for this study. The study is a phenomenological research design and data was collected through semi-structured interviews with 23 participants. Participants were recruited from three elementary schools, two middle schools, and one K-8 school within one school district in North Carolina. Participants taught in grades four through eight and had their own value-added data from the previous school year, since it seemed likely that teachers with their own value-added data (and not school-wide data) to populate Standard Six would have more input about how they use the information to reflect upon their own practices. Interviews were transcribed from audio recordings. Member checking and peer review methods were used. This chapter concluded with a review of my initial subjectivity and how I addressed this, how I ensured trustworthiness of the study, the benefits and risks involved, and limitations of the study.

As I moved forward with this study, I was hoping to enhance my understanding of the teacher's point of view as it relates to the new evaluation system in North Carolina and how I, as an administrator, could improve my own skills when providing orientation for new teachers to the evaluation process. I also hoped to become more aware of how teachers perceive the evaluation process and what benefits, if any, it has on their own

practices and professional growth. Finally, I hoped to be able to offer insight into the teacher evaluation process in North Carolina and how we can continue to improve upon the process for the purpose of ensuring quality teachers for all students.

CHAPTER II

REVIEW OF THE LITERATURE

The Culture of Accountability in Education

Throughout the majority of the nation's history, many political leaders and citizens believed that the United States provided the best opportunities in the world. Then, in 1957, the Soviet Union launched *Sputnik* and the commencement of America's race with the Soviet Union to improve national defense programs truly began. Until that time, the vast majority of Americans likely never thought another country could compare to the successes of the U.S. with any measurable criteria, and certainly not in terms of educating its citizens. However, in 1983, *A Nation at Risk* was published and concern grew even more over whether or not the U.S. public schools were truly providing students with a superior education (Spring, 2011).

Recently, the reality is that numerous countries and educational systems around the world are matching U.S. performance in educational outcomes and, in fact, surpassing them in some cases. Our educational system has made negligible improvements since the seminal *Nation at Risk* report, despite the fact that we have doubled educational spending (Klein, 2011; Mourshed & Barber, 2007). An examination of national and international reports on educational performance suggests that the United States of America does not seem to be producing equitable outcomes for all students, nor is the United States the top performer in our global community in educating its youth. According to the National

Assessment of Educational Progress (NAEP), 52% of high school seniors were only at a basic level of writing proficiency. Also, only 75% of students in U.S. public schools graduate from high school within four years (Chapman, Laird, Ifill, & KewalRamani, 2011). At this rate, one in every four young adults does not have a diploma to advance to post-secondary education, which is increasingly becoming the likeliest avenue for young adults to compete in a global job market.

International assessments that compare U.S. students to other nations' students indicate that while many students are performing well, others are not even at a basic skill level in reading, math, and science. For example, Progress in International Reading Literacy Study (PIRLS) is an assessment of reading comprehension conducted by the International Association for the Evaluation of Educational Achievement (IEA). In 2006, 45 countries and other systems participated in the assessment. Compared to the other 44 participating countries and systems, the U.S. ranked 18th. The Russian Federation, Canada, Singapore, and the Netherlands were among countries that outperformed the U.S. in 2006. The 2011 administration of the PIRLS resulted in a considerable improvement in performance for the U.S., ranking sixth behind Hong Kong, Russian Federation, Finland, Northern Ireland, and Singapore.

The Trends in International Mathematics and Science Study (TIMSS) is an assessment developed and implemented by the IEA as well. It was developed by IEA to measure the knowledge of fourth- and eighth- grade students in the areas of mathematics and science. In the U.S., results are published by the National Center for Education Statistics (Gonzalez, Williams, Jocelyn, Roey, Kastberg, & Brenwald, 2008). In 2007, the

U.S. ranked 11th among all participating countries and systems on the mathematics assessment given to fourth graders and 9th among all countries and systems on the mathematics assessment given to eighth graders. Countries such as Singapore and Japan consistently obtained higher scale scores on average than the U.S. in all domains. In 2011, the U.S. again ranked 11th with the fourth grade results and 9th with eight grade results. On the science portion of the TIMSS in 2007, the U.S. ranked 8th among all of the participating countries and systems comparing the performance of fourth grade students. Eighth-grade students in the U.S. ranked 11th among other countries and systems. Again, Singapore consistently scored significantly higher on average on all domains compared to the U.S. Japan scored significantly higher than the U.S. on all domains but one. In 2011, the U.S. fourth graders ranked 7th and the eighth graders ranked 10^{th}.

The Program of International Student Assessment (PISA) is an assessment conducted every three years internationally to assess how well 15-year-old students apply their knowledge in reading, science, and math within real-world contexts. In 2009, PISA was administered in 34 countries that are members of the Organization for Economic Cooperation and Development (OECD), and over 30 countries and systems not members of the OECD. On the PISA 2009 reading assessment, the United States ranked 12th among OECD countries. Out of all countries and systems, 15 scored better on average than the United States. On the PISA 2009 math assessment, the U.S. ranked 25th among OECD countries. Out of all countries and systems, 30 scored better than the U.S. Countries that have consistently scored better than the U.S. in reading literacy since 2009

include, but are not limited to, Australia, Canada, Finland, Japan, Korea, New Zealand, Norway, and Singapore.

In addition to public education's outcomes within an international context, achievement gaps nationally between schools and between ethnic and racial groups tend to further highlight the fact that achievement outcomes for many students in the United States are lacking. For example, some schools are outperforming other schools, even when they are serving the same populations of students, such as inner-city charter schools that outperform adjacent schools in the same neighborhoods (Morgan, 2012). Also, there are significant and persistent gaps across the nation between economically advantaged students and their counterparts as well as gaps between minority students and Caucasian students (Morgan, 2012).

In summary, many students in the United States are not performing in reading, writing, math, or science at a level of proficiency required to be successful and competitive citizens in the global work force. Because of these discrepancies in student outcomes, concern exists that differences in outcomes are due to differences in teacher efficacy (Thomas, Wingert, Conant, & Register, 2010; Waddell, 2012). Many politicians, policymakers, and educational philanthropists believe that one possible way our country will begin to make improvements is through the concept of accountability entering the profession of education.

Accountability for educators based on student growth and performance is a new concept in our history of public schooling. For decades, teachers have enjoyed tenure and the support of teachers' unions, protecting them from being removed from the classroom

(Morgan, 2012; Stover, 2011; Thomas et al., 2010; Waddell, 2012). Recently, many have begun to believe that student achievement data should be part of teacher evaluations (Sawchuck, 2012). The *No Child Left Behind Act of 2001* contributed greatly to the growing emphasis on teacher accountability with its mandates related to standardized testing and schools ultimately being forced to guarantee student outcomes. Under the Obama administration, the federal government's push for teacher quality has further played an important role in this growing sentiment. The *Race to the Top* incentive program has encouraged numerous states, beginning with Delaware and Tennessee, and including North Carolina, to increase teacher accountability by attaching student growth data to teacher evaluation (Schachter, 2010; U.S. Department of Education, 2009). In addition, both the RAND Corporation and the Bill and Melinda Gates Foundation, two of the most prominent research and non-profit organizations, support the use of student growth data being used to measure teacher effectiveness, though they do believe it should not be used as the sole measure of teacher effectiveness (Duffrin, 2011).

The increasing demand for teacher accountability has become so strong that some districts in some states have even publicized teachers' effect on student growth. For example, in 2010, the Los Angeles school district released the value-added data of its teachers. Then, in 2012, New York City's Department of Education released the value-added data of its teachers to the media (Barile, 2013). Nineteen states allow student performance data to be released to the public due to open records laws, even though many proponents of using student outcome data as part of teacher evaluations argue against making such data public (Sawchuck, 2012). The public dissemination further

fuels the current debate related to teacher accountability and complicates decisions surrounding the issue.

Even in areas where teacher effects on student growth are not made public, there are still high-stakes decisions being made about teachers. For example, Washington, D.C. dismissed hundreds of teachers under former school chancellor Michelle Rhee and has continued to do so even after her groundbreaking term (Winters, 2012). Teachers in other areas, including states such as Nevada, Colorado and Tennessee, are changed back to probationary status if they have poor student growth data for two years in a row (Winters, 2012). In Colorado and Tennessee, as well as in Louisiana, at least half of the teacher evaluation instrument is weighted with student performance data (Green, Baker, & Oluwole, 2012). In Florida and Michigan, parents must be notified if their child's teacher consistently performs poorly on the teacher evaluation instrument (Sawchuck, 2012).

In addition, in some states, unions are now prohibited from bargaining on anything related to teacher evaluation, including specifics related to content and process (Paige, 2013). This speaks to the increasing power of teacher accountability demands, since teacher unions have long possessed significant bargaining control over teacher personnel issues. One such state is Wisconsin, where unions are now unable to negotiate any portion of their teacher evaluation process. In fact, Wisconsin has added value-added data as a component of teacher evaluation. This is also true in Michigan and Florida. In Michigan, reduction in force (RIF) policies now include value-added teacher data as the major factor in determining who is affected, instead of seniority, which is the method favored by unions (Paige, 2013). In Colorado, the American Federation of Teachers

affiliate actually endorsed a bill that altered teacher evaluation in the state to include student growth outcomes (Goldstein, 2011). Once receiving Democratic support, unions are ever increasingly finding themselves alone in the argument to protect teacher rights at what the public sees as at the expense of student success (Stover, 2011; Toch, 2010/2011).

The Importance of Teacher Quality

Ever since the seminal *Coleman Report* published in 1966, researchers have been increasingly determining the impact of teacher characteristics and teacher quality on student outcomes. Today, the importance of teacher effectiveness is well established from years of research (Hanushek, 1992; Looney, 2011; Sanders & Rivers, 1996). Hanushek and Rivkin (2006) for example, determined that students assigned to the lowest performing teachers only make a half year's growth in reading, while students assigned to the highest performing teachers make one and a half years' growth in reading. Klein (2011) ponders the cumulative effects of having the best teachers over the expanse of kindergarten through twelfth grade and claims "the impact would be seismic" (p. 73). Winters (2012) agrees and states that the effects are "sustained throughout the child's life" (p. 76).

Klein (2011) quotes President Obama saying in 2008 that the most important factor in students' achievement is who the teacher is in the classroom. To illustrate this point, he describes the Harlem Success Academy in New York City, which accepts students by lottery. These students are primarily minority students and students of low socioeconomic status and 88% of them are proficient in reading and 95% of them are

proficient in math. Numerous other schools in the same area serving students of the same demographics have approximately 30-40% of the students achieving proficiency in reading and math. In addition, he points out that the Harlem Success Academy "now performs at the same level as the gifted-and-talented schools in New York City" (p. 73). Klein (2011) argues that the factor that distinguishes the Harlem Success Academy from other, nearby schools is the placement of a highly effective teacher in each classroom. Hanushek (2007) agrees that the achievement gap between Caucasian and African-American students is largely due to the inequitable access to high-quality teachers faced by minority students and students of low socioeconomic status.

A pivotal study conducted by Sanders and Rivers (1996), found that students who had three consecutive teachers having high value-added data gained 50 percentile points on an achievement test than did students who had three consecutive teachers who fell within the average range. A study conducted by Stronge et al. (2011) found similar results. The students in their study who were taught by teachers falling in the highest quartile of value-added data performed over 30 percentile points higher in both reading and math than students taught by teachers in the lowest quartile of value-added data.

Internationally, there is further evidence of the importance of teacher quality in student outcomes. Mourshed and Barber (2007) found that the highest performing school systems in the world recruit teachers from the top third of graduates, whereas the lowest performing school systems recruit the majority of their teachers from the bottom third of graduates. In Singapore, the best teachers teach the students with the most academic difficulties and the result is that Singapore has among the best scores on international

tests (Morgan, 2012). Darling-Hammond (2008) argues that leading countries such as Singapore, Japan, Finland, and Sweden ensure that teachers are highly trained and accessible to the least advantaged students. In contrast, almost half of the U.S. teaching force come from the bottom third of performers on the SAT and many of these teachers end up teaching in predominantly low-income schools, and the result is a significant achievement gap between economically disadvantaged students compared to their peers (Kristoff, 2011; Mourshed & Barber, 2007).

Hanushek (2007) contends that the economic health of our country is directly correlated to student achievement and that teacher quality is the key to improving student achievement. Mourshed and Barber (2007) argue that raising teacher quality to a comparable level of leading countries in terms of student achievement would raise our gross domestic product by at least two trillion dollars. Payne and Schleiler (2011) suggest that a Stanford University study by Hanushek estimates a $41 trillion increase for the United States economy if we raised student achievement to the level of current leading countries. A failure to do so will lead to a "gloomy future" for the nation, Morgan (2012) argues (p. 296).

A gloomy future also awaits individual students who do not graduate from high school with necessary skills to be successful in a postsecondary program. In the current job market, even blue-collar jobs are requiring more advanced skills due to advances in technology and the current economy makes it significantly difficult to survive on minimal wages (Jacob, 2007).

Teacher Quality Research

Although the importance of teacher quality to student outcomes is not disputable, Hanushek (2007) argues that common measures of teacher quality are "largely uncorrelated with true quality" (p. 574). There is some debate on whether teacher effectiveness should be measured by a teacher's qualifications, their instructional practices and delivery, their effects on student achievement, or a combination of these (Lewis et al., 1999; Stronge et al., 2011). Stronge (2002) provides a framework for teacher effectiveness based on a meta-analysis of the research on the subject. This conceptual framework includes three dimensions of professional skill: instruction, assessment, and the learning environment. Personal qualities comprise the fourth area, or dimension, of teacher effectiveness. Stronge's conceptual framework is incorporated into the working conceptual framework illustrated in Figure 1.

The first dimension, Instructional Delivery, has multiple components. This dimension includes the teacher's effectiveness in direct instruction, individualized or differentiated instruction, and the use of a variety of instructional strategies. Effective teachers maintain students' focus on learning and they promote critical thinking skills. Effective teachers are articulate, provide detailed directions, and explain the content clearly, all of which is referred to as instructional clarity. Effective teachers also understand the content they teach and transfer that knowledge to students in a way that allows them to conceptualize the relationships between content matter within and between subjects. This is referred to as instructional complexity. Also, effective teachers communicate high expectations to their students through what they require of their

students and by holding them accountable for work completion. They use technology during instruction and explore a variety of technology resources.

The second dimension of Stronge's conceptual framework is Student Assessment. Effective teachers use formal as well as informal assessments to monitor student learning and to provide specific and timely feedback to students. They also use assessment information to adjust instructional plans based on their students' performance.

Stronge's third dimension is the Learning Environment. Effective teachers are organized and establish routines for productivity. They also establish positive relationships with their students based on fairness and respect. They set clear expectations and reinforce them regularly.

Personal Qualities is the final dimension of Stronge's conceptual framework. The most effective teachers are enthusiastic about what they do. They make positive connections with their students and encourage them to take responsibility for their own learning. They are also highly reflective.

Stronge et al. (2011) found in a study that examined over 4,600 students' performance in reading and math that teachers' experience levels had no relationship with student achievement gains. This lack of relationship has been found in other studies as well (Jacob, 2007). However, teachers' value-added data were strongly correlated with student achievement gains. Students taught by teachers in the highest quartile scored more than 30 percentile points better in both reading and in math than students taught by teachers in the lowest quartile (Stronge et al., 2011).

When comparing the teachers' instructional delivery, teachers in the highest quartile had fewer class disruptions, they had higher expectations for student behavior, established more routines and procedures, had more positive interactions with students, and required students to take a greater degree of responsibility in the classroom (Stronge et al., 2011). A study by Rockoff (2004) of approximately 10,000 students and 300 teachers found that differences in teacher quality, as judged by multiple observations, were correlated with significant differences in student outcomes. It is this difference in quality that value-added measures hope to capture.

Hanushek (1992) described the difference between the best and worst teachers based on value-added measures as equivalent to about 1.5 years of growth for students. In other words, the students taught by the teachers with the best value-added data will achieve about 1.5 years of growth during one school year while the students taught by the teachers with the worst value-added data will only achieve about a half year of growth within a school year.

While value-added measures of teachers are highly predictive of student performance, teacher characteristics that currently drive policy and practice seem to have no predictive value of student achievement. Characteristics such as experience level, traditional or alternate certification, or level of education (such as the attainment of advanced degrees) have no predictive value of student outcomes (Hanushek, 2007; Jacob, 2007; Stronge et al., 2011). Hanushek (2007) succinctly states, "The importance of teacher quality for determining student outcomes is now well established. At the same

time, the translation of what is known about teacher quality into effective policy is far from being institutionalized" (p. 574).

History of Teacher Evaluation

For much of the teaching profession's history, teachers were not evaluated *per se* based on their teaching performance but often by unfair and unrelated reasons to their teaching and so teacher tenure was used to protect teachers from this unjust practice (Chelsey, 2011; Thomas et al., 2010). Today, teacher tenure is often argued as a way to protect incompetent teachers from being dismissed. For this reason, discussion about tenure sometimes overlaps into discussion about teacher evaluation. However, these are two very separate issues and teacher evaluation is more relevant to understanding some of the more problematic issues with ensuring teacher quality, as the current study will demonstrate.

For well over a hundred years, teacher evaluations have rated over 99% of teachers as effective (Kimball, 2002; Papay, 2012; Winters, 2012). Almost 50 years ago, Perry (1977) described the teacher evaluation process as "careless to the point of being desultory if not nonexistent" (p. 185). This "rubber stamp" phenomenon persists even when student outcomes do not indicate that effective teaching and learning has occurred, as determined by reading proficiency. For example, Winters (2012) explains that over 50% of fourth grade students were not proficient in reading, yet less than 1% of their teachers were rated as having unsatisfactory performance. Likewise, in Houston, over 40% of fourth graders were not proficient in reading, yet only 3% of their teachers were rated as having unsatisfactory performance. Winters goes on to explain that the tendency

to rate the vast majority of teachers as effective is in contrast to empirical evidence and "simple common sense" (p. 63). He and Duffrin (2011) cite a 2009 survey (conducted by The New Teacher Project) in four large school districts that found over 40% of teachers believe they work with at least one other teacher who is so ineffective they should not be teaching. He also cites a 2007 survey of veteran teachers that found over half of the teachers felt that they received inflated performance ratings. Thomas and Wingert (2010) cite that even the American Federation of Teachers' former president, Randi Weingarten, acknowledged that there is a difference between the number of incompetent teachers and the number who actually get dismissed.

Although one purpose of teacher evaluation is to measure teacher effectiveness, another primary purpose of teacher evaluation is to stimulate the professional development of teachers (Feeney, 2007; Looney, 2011; Papay, 2012). However, many teacher evaluation systems do neither of these effectively (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). Looney (2011) urges educational systems to "find an appropriate balance" between the purposes of quality assurance and teacher development (p. 440). Currently, under President Obama's presidency, the federal government has encouraged states to incorporate value-added data into teacher evaluations through its Race to the Top funding as a way to improve the discernment of effective and ineffective teachers. Currently, many states and large, urban school districts are using it to calculate part of teachers' evaluations (Goldstein, 2011; Winters, 2012). North Carolina is one of these states currently using value-added data in teacher evaluations. However, Di Carlo (2012) argues that "virtually no empirical evidence as to

whether using value-added or other growth models—the types of models being used vary from state to state—in high-stakes evaluations can improve teacher performance or student outcomes" (p. 38). He further contends that it will likely be "several years before there is solid initial evidence on whether and how the various new evaluation systems work in practice" (p. 38).

Defining Value-Added Data (VAD)

Value-added data, subsequently referred to as VAD, compares how a student has done on a standardized test compared to how a student was predicted to perform based on factors not under the teacher's control, such as previous test performance, socioeconomic status, English proficiency, class size, race, and gender (Barile, 2013; Di Carlo, 2012; Doran & Fleischman, 2005; Duffrin, 2011; Green, Baker, & Oluwole, 2012; Papay, 2012; Winters, 2012). This is accomplished through a statistical procedure known as multiple regression (Winters, 2012). The difference between how much growth the student is expected to make and how much growth the student actually makes is considered to be the teacher's effect on the student. Once this is done for every student in a teacher's class, the result is an overall measure of the teacher's positive or negative value on student achievement, which is referred to as value-added data of the teacher (Barile, 2013; Di Carlo, 2012; Doran & Fleischman, 2005; Duffrin, 2011; Green, Baker, & Oluwole, 2012; Papay, 2012; Winters, 2012). The mean value-added score is set to zero, so if a teacher's value-added score is negative, his or her effect on students was below average compared to other teachers in the district or state (depending on what the specific school system chooses to use as its pool of comparison teachers). If the teacher's value-added score is

positive, his or her effect on students was above average compared to other teachers in the district or state (Winters, 2012).

Over time, a teacher who consistently obtains higher values would be considered a highly effective teacher, whereas a teacher who consistently obtains negative values would be considered an ineffective teacher (Hanushek, 2007). However, proponents of its use in teacher evaluations urge that it be used in combination with classroom observations and not as a sole measure of a teacher's performance (Duffrin, 2011; Winters, 2012). In addition, proponents urge that it be used only after at least three years of data can be collected so that a sufficient number of students' scores are used in the calculation of a teacher's value-added data and so that a pattern of performance can be considered (Winters, 2012). Value-added data is better than standard growth data because growth data simply represents the difference between actual students' scores and their expected scores based on their own previous performance and the average rate of growth, regardless of other factors (Goldstein, 2011).

Arguments in Favor of Using VAD in Teacher Evaluations

William Sanders and Robert McLean originally used a statistical methodology while at the University of Tennessee in the early 1980s that became known as a "value-added" formula for determining teacher effectiveness because of their claims that the statistical methodology could isolate a teacher's effect on student outcomes (Sanders & Horn, 1994). What Sanders and McLean found was that significant differences were evident between groups of students taught by different teachers, that evaluations by teachers' supervisors were strongly correlated with the teacher's "effect" on students, and

that student growth was not related to students' ability levels or proficiency levels (Sanders & Horn, 1994). Since that time, value-added data has become a widespread phenomenon in education and is supposed to provide an alternative or supplemental information to what is sometimes referred to as a "rubber-stamp approach" to the teacher evaluation process, meaning that very few teachers are rated poorly on current evaluation methods that utilize observations measured with rating scales (Winters, 2012).

Value-added data provide evidence that students have learned the material taught to them and, thus, is an objective measure of how effective the teacher has taught (Duffrin, 2011; Sanders & Horn, 1998). It is typically consistent with principal ratings based on classroom evaluations, which means that this data can validate subjective ratings during the teacher evaluation process by providing objective evidence of the teacher's performance (Duffrin, 2011; Goldstein, 2011; Sanders & Horn, 1998; Winters, 2012). In fact, observable qualities that tend to differentiate between effective and ineffective teachers correspond with value-added data. For example, teachers in the bottom quartile of value-added measures have disruptive behaviors occur in their classrooms three times as often as teachers performing in the top quartile (Stronge et al., 2011). These researchers also found that teachers in the top quartile had more routines and procedures, were more organized and prepared, communicated higher expectations, and had more positive interactions with students. Observers in this study had no knowledge of teachers' value-added data, so their observations could not be influenced by the knowledge.

Indeed, some researchers argue that teachers' value-added data are much stronger predictors of how students will perform than other teacher factors that we currently emphasize, including years of experience and level of pay, or student factors that we emphasize as predictors, including socioeconomic status (Sanders and Horn, 1994; Stronge et al., 2011). Also, because value-added measures specify the amount of growth a teacher's students have experienced, they are much more valid than judging a teacher's performance based on the number of students identified as proficient because qualifiers such as "proficient" are arbitrarily set by each state and vary significantly (Good, Wiley, & Sabers, 2010; Sanders & Horn, 1994).

According to Olson and Sabers (2008), North Carolina had the lowest standard of all states for deeming a student as proficient in reading and, thus, 88% of students were identified as proficient. In contrast, South Carolina had the highest standard for deeming a student as proficient in reading and, therefore, only 30% of the students were identified as proficient. Since it does not make sense to evaluate teachers without any consideration at all of student outcomes, the fairest way to include student performance in teacher evaluations is through the use of value-added data.

Winters (2012) examined the value-added data from 2007 linked to teachers in Florida and found that teachers with the poorest value-added data were teachers administrators wanted to remove from classrooms. He also found that these teachers continued to underperform two years later. Thus, Winters argues that value-added data can be a valuable tool to identify the worst performing teachers in our schools. Hanushek (2007) agrees. He found that teachers with high value-added scores typically achieve 1.5

years of growth with their students, whereas teachers with negative value-added scores often achieve only half of a year of growth with their students. Gordan, Kane, and Staiger (2006) estimate that denying tenure to the bottom quarter of new teachers would result in significant improvement in overall student achievement, arguing that replacing the worst-performing teachers with average-performing teachers, or teachers with moderate value-added scores, would create a gain in achievement.

Goldhaber and Hansen (2008) found that value-added data of teachers in the first three years of experience accurately predicts how well subsequent students will perform in those teachers' classrooms. They also found that value-added data more accurately predicts teacher performance in later years than do factors such as teachers' level of experience or attainment of advanced degrees. Galley (2011) agrees that value-added data measurement is an objective tool to determine who is underperforming in the classroom, which can lead to the removal of ineffective teachers when part of a teacher evaluation system. For example, New York City schools have used value-added data to deny teachers tenure since 2010, which they feel is helping to improve teacher quality (Duffrin, 2011). The Houston Independent School District, which has used value-added data in teacher evaluation since 2007, has decreased the retention of its lowest performing teachers from 13% to 2% since that time (Duffrin, 2011).

Winters (2012) argues that quantifiable data in education has drawn more quantitative researchers into educational research. He points out that many of these quantitative researchers are economists holding influential federal government positions or university positions. Such individuals have the potential to impact the field of

education in positive ways by increasing our ability to identify differences in teacher quality and influencing reform efforts that can improve teacher quality.

One such economist is Eric Hanushek from Stanford University. This proponent of value-added data does not encourage its use in isolation. Rather, he believes that value-added measures should be used in conjunction with other means of teacher evaluation (Hanushek, 2007; Winters, 2012). He further argues that opposition to the use of value-added measures in teacher evaluations are primarily the result of more concern over the job security of teachers, irrelevant of actual performance, and less concern about students' access to an equitable education. Nevertheless, even he acknowledges that VAD has its limitations. Specifically, VAD tends to be able to distinguish between the most and least effective teachers but is less able to distinguish between teachers in the middle area of performance, which is where the majority of teachers perform. However, Hanushek (2011) argues that identifying and eliminating the worst performing teachers would have a significantly positive effect on the quality of the teaching workforce that would translate into substantial improvements on the long-term outcomes of countless students entering the global marketplace.

Arguments against the Use of VAD in Teacher Evaluations

Sean Corcoran from New York University is also an economist, like Hanushek, but Corcoran has more reservations about the use of VAD. He points out that VAD formulas lack transparency and have limited face validity to practitioners which can hinder their practical utility in terms of teacher reflection on practice (Jennings &

Corcoran, 2009). He cautions that VAD be used "only as a part of a well-rounded evaluation system" due to its limitations (Corcoran & Goldhaber, 2013, p. 425).

There is more than one statistical model to determine a teacher's value-added data, which is one reason why its use is controversial; a teacher's value-added data can look differently depending on which formula is used (Darling-Hammond et al., 2012; Di Carlo, 2012; Green et al., 2012). Also, some researchers argue that value-added formulas are based on the assumption that students are randomly assigned to classes, but that is often not the case since students are often placed strategically in classes at many schools (Darling-Hammond et al., 2012; Good et al., 2010; Green et al., 2012). Papay (2012) argues that this is perhaps the greatest weakness of VAD.

Also, if too few student scores are used to calculate a teacher's value-added data, then there is an increase in the possibility that the results are due to chance factors (Di Carlo, 2012; Duffrin, 2011). Green et al. (2012) argue that value-added measures are too statistically weak and have too much error due to extraneous factors, which is why multiple years of data are needed. Only a third of teachers in the highest or lowest quintiles actually stay in those quintiles from one year to the next (Darling-Hammond et al., 2012; Green et al., 2012). Green et al. (2012) also argue that even after three years of data, the error rate can still be as high as 25%.

Although value-added measures attempt to control for factors such as low socio-economic status and other factors outside of a teacher's control, these measures may not be able to control for every factor outside of a teacher's control (Doran & Fleischman, 2005; Goldstein, 2011; Good et al., 2010; Paige, 2013; Winters, 2012). For example,

Jacob (2007) explains that urban schools have a set of unique factors that may not be accounted for in value-added data, particularly when urban teachers are compared to suburban or rural teachers. He describes the concentrated existence of over 120 languages spoken in New York City, which limits the amount of communication between home and school, as well as the high rates of mobility and the lack of social capital among many urban neighborhoods which places children at even more of a disadvantage. Social capital refers to the connections between individuals in a community that help monitor and influence each other in positive ways, thus providing an informal support or network (Jacob, 2007).

Jacob (2007) also describes the inequitable resources between schools and the poor tax bases in many urban settings that create a disparity between urban schools and other schools that have strong local funding. Goldstein (2011) cites factors such as parental involvement, tutoring, student motivational factors, and life events as factors that cannot be accounted for through the use of value-added data. Can value-added measures possibly take all of these factors into account when comparing teacher performance? Galley (2011) agrees that value-added data should only be used in a local context, where supervisors are aware of other factors relevant to a teacher's value-added results.

Some who oppose the use of value-added data in teacher evaluations do so not necessarily because of the nature of value-added data itself, but because of the standardized tests used to calculate the data, which some describe as "unsophisticated" tests (Goldstein, 2011, p. 18). Opponents to the use of standardized test results in teacher evaluations claim that the tests do not adequately measure changes in students'

knowledge over time and that the tests were not designed to measure growth or to be used in teacher evaluations (Galley, 2011; Gitomer, 2011; Good et al., 2010). This can particularly be true of tests that have what is referred to as a "ceiling effect," which is when the test can not sufficiently distinguish between high levels of performance due to students achieving perfect scores (Papay, 2012). Also, determining which test to use can have a drastic impact on a teacher's VAD. For example, two tests measuring the same content area in the same grade during the same year can yield significantly different VAD results on the same teacher (Corcoran, Jennings, & Beveridge, 2011).

Goldstein (2011) highlights the fact that at least 10 years of data is required to reduce the error rate to 12 percent or less. Good et al. (2010) argue that curriculum content spans over multiple grades and, therefore, a test is not just measuring the current teacher's effectiveness. How well or poorly a student does on a test in any given year is impacted by how effective or ineffective the previous teacher was in teaching the necessary prerequisite information. Thus, if a student does not learn the information from a grade level one year and he or she is still placed in the following grade, the child's beginning score places the teacher at a significant disadvantage.

Galley (2011) highlights additional flaws with value-added data. She argues that the use of value-added data in teacher evaluations will likely narrow the curriculum even more so than that which has already happened under *No Child Left Behind*. Barile (2013) and Larsen (2005) agree that the use of value-added data will cause teachers to "teach to the test" and discourage creativity in the classroom. Galley (2011) and Marshall (2012) also argue that too much emphasis on value-added data in teacher evaluations could

possibly delay the termination of a teacher, since several years of data is necessary to increase the validity. A teacher could argue, therefore, that not enough time has been given before making a personnel decision of such magnitude. Finally, Galley (2011) also warns that the heated debate over value-added data ignores equally, if not more, important reform efforts "such as strengthening the curriculum, encouraging teacher collaboration, and improving instructional leadership" (p. 45).

Jacob (2007) contends that because only reading and math teachers in third through eighth grades have value-added data, the use of such is "only a partial solution to the issue of ineffective teachers" (p.147). In fact, Green et al. (2012) estimates that value-added data can be calculated for only 20% of teachers. Papay (2012) agrees that less than a third of teachers have their own VAD. Winters (2012) adds that it is not sufficient to utilize a tool that does not provide us with information about a large number of our teachers or to solely rely on a tool that does not give information about what precisely a teacher is doing correctly or incorrectly. O'Donovan (2011) argues that proponents of value-added measures erroneously assume that teacher evaluation will improve teacher performance or that value-added data will lead to the replacement of bad teachers with better ones. He argues that this is highly unlikely since "schools would use the same inefficient process that placed the ineffective teachers in the first place to find a replacement" (p. 73).

Di Carlo (2012) contends that the debate about value-added data should not be whether or not to use it, but how to use it. He argues that any single measure of teacher quality in isolation will lead to error and that observations of teacher performance also

result in variance and imprecision over time. He suggests that the key to utilizing value-added data is to use it as but one of several ways to identify our teachers most in need of support or replacement. Others agree (Hanushek and Rivkin, 2010; Winters, 2012). However, even if value-added measures are used in conjunction with classroom observations to evaluate teachers, there are still many teachers who will not have their own value-added data. While the suggestion exists to create standardized tests for every subject and every grade, Galley (2011) points out that this may not be the wisest use of educational funds, particularly when there are multiple value-added models, none of which can take into account all factors that influence children's academic growth.

Some districts, such as Los Angeles and New York City, have published teachers' value-added data. The fear of publication is another factor opponents argue against the use of value-added data in teacher evaluations (Barile, 2013). Other components of a teacher evaluation are subjective and, therefore, automatically exempt from public disclosure. While value-added data is numerical, Barile (2013) argues that it is erroneous "to assume that a number, simply because it is the product of a mathematical formula, is necessarily an objective fact" (p. 136). Papay (2012), too, explains that, because VAD is numerical, it is sometimes perceived, particularly among policymakers, as being more unbiased as standards-based evaluation approaches. However, like Barile (2013), he argues that this is misleading.

Papay (2012) identifies three criteria by which both VAD models and standards-based evaluation models should be judged. Specifically, he argues that any model should be 1) unbiased, 2) reliable, and 3) valid. He has found that both standards-based

evaluation and VAD models suffer from deficits in all three areas. However, Kane, Taylor, Tyler, and Wooten (2011) have found that standards-based evaluation, when done with fidelity, can predict student test performance better than VAD.

Related Law and Case Law

The most fundamental protection of teachers' rights related to job retention is the Fourteenth Amendment of the Constitution that states every person is entitled to due process before the loss of life, liberty, or property. There are two types of due process: procedural due process and substantive due process. Procedural due process is the right to a fair process before the decision is made to take away a person's life, liberty, or property interest. Substantive due process imposes limitations upon processes or procedures (Green et al., 2012). State tenure laws also protect teachers against "arbitrary or capricious" dismissal decisions, as cited in *Frank Cona v. Avondale School District* (Darden, 2013).

A teacher's liberty interest would be his or her personal standing in the community or his or her ability to obtain other employment. Case law such as *St. Louis Teachers Union, Local 420 v. Board of Education of St. Louis* determined that a teacher being evaluated against a standard of performance is not an example of loss of liberty. However, Green et al. (2012) argue that someone terminated based on VAD may be able to establish that a property interest has been lost in states where tenure laws exist, particularly due to the fact that VAD has such a high potential for error. He cites the second factor under *Mathews v. Eldridge*, in which case the Supreme Court held that a

risk of erroneous deprivation of property through the procedures used was one of the determining factors in deciding whether or not procedural due process was violated.

Green et al. (2012) also contend that challenges against VAD could be made under Title VII of the Civil Rights Act because of inherent bias of standardized tests against African American students and also due to the fact that African American teachers are more likely to work in schools of lower income, African American students. These researchers say that the case of Title VII violation, again, is more likely in states where at least 50% of the teacher evaluation is based on VAD. The violation would most likely be found under a disparate impact, which occurs when a policy seems fair, but has a discriminatory effect.

In addition to teachers' concern over their right to retain their jobs in the face of increased accountability, this increased focus, combined with the use of VAD, may also place teachers at risk of educational malpractice (DeMitchell, DeMitchell, & Gagnon, 2012). Under *Peter W. v. San Francisco Unified School District,* educational malpractice was deemed impossible due to the fact that a standard of care could not be established, since multiple factors are involved in a student's learning outcome and the teacher's effect could not be separated from the multitude of factors. DeMitchell et al. (2012) contend, however, that VAD may eventually be found in court to provide a standard of care, by which a teacher could conceivably be judged. DeMitchell et al. (2012) cite the court case in 1982, *Hunter v. Board of Education of Montgomery County,* in which the judge stated that determining a standard of care in the profession of teaching is appropriate and that educational malpractice is feasible.

North Carolina Educator Evaluation System (NCEES)

The North Carolina Educator Evaluation System (NCEES) was fully implemented in North Carolina during the 2010-2011 school year. Though some school districts have two evaluation systems, one for teachers that have their own VAD and one for teachers who do not, North Carolina's teacher evaluation instrument does not differentiate between these two groups of teachers. All teachers are evaluated with the single formatted NCEES. The NCEES contains six standards. These standards are: 1) Teachers demonstrate leadership, 2) Teachers establish a respectful environment for a diverse population of students, 3) Teachers know the content they teach, 4) Teachers facilitate learning for their students, 5) Teachers reflect on their practice, and 6) Teachers contribute to the academic success of their students. Each of the first five standards is comprised of various elements of performance that demonstrate efficacy in the standard. The six standards and elements contained in each are listed in Table 1.

According to the North Carolina Teacher Evaluation Process (NCDPI, 2012b), some of the purposes of teacher evaluation include to serve as a measurement of performance, to serve as a guide for teachers as they improve their effectiveness, to serve as the basis for instructional improvement, and to guide professional development. If the evaluation process successfully accomplishes what it intends, then it is reasonable to conclude that teachers should be able to identify and speak to the standards that measure teacher effectiveness. They should also be able to provide examples of how the evaluation process has impacted their practices to lead to improvements in efficacy.

Table 1

North Carolina Educator Evaluation System Standards and Elements of Performance

Standard One: Teachers Demonstrate Leadership				
Teachers lead in their classroom.	Teachers demonstrate leadership in the school.	Teachers lead the teaching profession.	Teachers advocate for schools and students.	Teachers demonstrate high ethical standards.
Standard Two: Teachers Establish a Respectful Environment for a Diverse Population of Students				
Teachers provide an environment in which each child has a positive, nurturing relationship with caring adults.	Teachers embrace diversity in the school community and in the world.	Teachers treat students as individuals.	Teachers adapt their teaching for the benefit of students with special needs.	Teachers work collaboratively with the families and significant adults in the lives of their students.
Standard Three: Teachers Know the Content They Teach				
Teachers align their instruction with the North Carolina Standard Course of Study.	Teachers know the content appropriate to their teaching specialty.	Teachers recognize the interconnectedness of content areas/disciplines.	Teachers make instruction relevant to students.	
Standard Four: Teachers Facilitate Learning for Their Students				
Teachers know the ways in which learning takes place, and they know the appropriate levels of intellectual, physical, social, and emotional development of their students.	Teachers plan instruction appropriate for their students.	Teachers use a variety of instructional methods.	Teachers integrate and utilize technology in their instruction.	
Teachers help students develop critical-thinking and problem-solving skills.	Teachers help students work in teams and develop leadership qualities.	Teachers communicate effectively.	Teachers use a variety of methods to assess what each student has learned.	

Table 1

(Cont.)

Standard Five: Teachers Reflect on Their Practice		
Teachers analyze student learning.	Teachers link professional growth to their professional goals.	Teachers function effectively in a complex, dynamic environment.
Standard Six: Teachers Contribute to the Academic Success of Students		
A teacher's rating is determined by a student growth value calculated by the statewide growth model and places a teacher into one of three rating categories: "Does not meet expected growth," "Meets expected growth," or "Exceeds expected growth."		

According to Shakman et al. (2012), North Carolina was one of only two states that had implemented a statewide performance-based teacher evaluation tool that covered all of the teaching standards identified by the Interstate Teacher Assessment and Support Consortium (InTASC) by the 2010-2011 school year, with the other state being Texas. Since the NCEES is so closely aligned with the InTASC standards, it is, again, reasonable to conclude that teachers should be able to speak to the standards on which they are evaluated.

Summary

This chapter provided a summary of major historical events that contributed to the growing culture of accountability in public education. Such events included the space race with the Soviet Union in the 20th century and the launching of *Sputnik* in the 1950s. The seminal publication *A Nation at Risk* also contributed greatly to the increasingly critical public toward education. Legislation, including the *No Child Left Behind Act of 2001* under the Bush administration and *Race to the Top* under the Obama administration

solidified the political and public awareness of public education outcomes in the U.S. and firmly established the concept of accountability in education as both a bipartisan and pervasive expectation among the public for America's schools. This growing expectation for accountability was further fueled by national and international reports from the NAEP, NCES, IEA, and OECD, as well as from assessment results on tests such as the PIRLS, TIMSS, and PISA.

This chapter also provided a review of the literature related to the importance of teacher quality, which is well established, and teacher quality research. Stronge's conceptual framework was described as a tool for understanding the dimensions of teacher quality. This chapter also provided an overview of the history of teacher evaluation, from the beginning days when teachers did not receive evaluations to the establishment of tenure as a way of protecting teachers from unfair dismissal. Performance-based evaluation of teachers emerged, followed by the concept of using student outcomes on standardized tests as a way to evaluate teachers. The use of data to evaluate teachers, known as value-added, is the newer method and arguments both in favor and against its use were discussed. This chapter provided a brief discussion of law and case law related to teacher evaluation. Finally, the NCEES was described and the dual purposes, ensuring teacher quality and promoting professional growth of teachers, of the NCEES were explained.

The next chapter will provide the purpose of the study, which is to explore teacher perceptions of performance accountability measures using the NCEES, which includes both standards-based evaluation as well as value-added data as a form of teacher

accountability. A justification for each research question will be provided. The research design will be explained as well. The study's participants, how they were recruited, and their work site demographics will be described. The study's instrumentation and procedures will be explained in detail. The next chapter will also explain how the study's data will be analyzed. My subjectivity as researcher will be presented. The trustworthiness of the study, as well as the limitations, will also be offered.

CHAPTER III

METHODOLOGY

Introduction

The purpose of this study was to explore teacher perceptions of performance accountability measures using the NCEES, which includes both standards-based evaluation as well as value-added data as a form of teacher accountability. I wanted to know if teachers perceived the components of the NCEES as fair measures of teacher performance. This study also examined teachers' experiences with the evaluation process and how their perceptions and experiences with the NCEES affected their classroom practices. Teachers' understanding of each standard within the NCEES was also examined because teachers' understanding of each standard likely affects the extent to which the instrument is effective in facilitating teacher reflection and growth. I also wanted to determine what suggestions teachers have to improve the evaluation process and the impact that the process of teacher evaluation and the NCEES has on teacher practices within the classroom.

Through this study, I sought to answer the overarching question: *What are teachers' experiences and perceptions about the evaluation process using the NCEES and how do these experiences and perceptions inform their practices?* According to the North Carolina State Board of Education, the purpose of the teacher evaluation process in North Carolina is to assess teacher performance but also to guide the creation of a plan

for ongoing professional growth for a teacher. Therefore, if the process is supposed to promote teacher growth, then teachers should be able to identify ways that the process has impacted their professional practices by referring to experiences they have had during the evaluation and feedback process. Papay (2012) highlights the fact that most research has focused on the identification of and differentiation between teacher effectiveness and ineffectiveness, but little attention has been given to the teacher evaluation process as a method of professional development.

I also sought to answer the following questions:

What do teachers understand about Standards One through Six of the NCEES?

If teachers do not have a thorough understanding of the standards or of the instrument, it is not likely to effectively facilitate reflection or have a significant impact on one's practices. The level of understanding that teachers can communicate has clear implications on how effectively teachers have been trained on the instrument.

What do teachers think about the fairness of the standards-based portion of the NCEES?

If teachers believe that standards-based evaluation is a fair measure of teacher performance, they will be more likely to perceive the ratings and feedback on the NCEES as valid and, thus, use the feedback to inform their practices. If they do not believe that standards-based evaluation is a fair measure of teacher performance, they will be more likely to perceive the ratings and feedback as an invalid reflection on their performance and less likely to use the feedback to inform their practices. Sedikides and Green (2004) and Leary (2007) describe human nature as being resistant to and even threatened by

feedback that is inconsistent with their positive self-concept. Since individuals desire to confirm their own existing view of self, any feedback that is inconsistent with that self-view must be perceived as having a degree of fairness and as being derived from someone who is oriented with their best interest in mind so that one feels less threatened by the feedback, thus being more open to self-reflection and self-concept clarity (Feys, Libbrecht, Anseel, & Lievens, 2008; Green, Sedikides, Pinter, & Van Tongeren, 2009).

What do teachers think about the fairness of value-added data as a component of the NCEES?

If teachers perceive VAD as a fair reflection of their effectiveness, they will be more likely to reflect upon the information to inform future practices. If they perceive VAD as unfair, they will be less likely to use the information to inform future practices. Again, Feys et al. (2008) and Green et al. (2009) found that negative information is poorly received from individuals unless it is from a trusted source.

What recommendations do teachers offer about each standard, the NCEES, and methods of teacher accountability for student performance?

I was very interested in knowing what ideas teachers have for teacher accountability and what suggestions they have to improve upon or enhance the current evaluation process in North Carolina. If teacher growth is a primary goal of the NCEES, then it is vital that we determine to what methods teachers will be responsive, if not the NCEES. As discussed in the previous chapters, since teacher quality has been determined to be a critical factor in student achievement growth, then increasing teacher quality has

significant implications for improving student achievement. Therefore, these are important questions to be answered.

Research Design

This study followed a qualitative research design using a phenomenological approach. The purpose of the phenomenological approach was to explore teachers' understandings and perceptions of accountability and teacher evaluation and, specifically of the NCEES. A phenomenological approach emphasizes the shared, lived experiences of a group of individuals (Lichtman, 2010). In this study, the shared experience was that of the teacher evaluation process in North Carolina using the North Carolina Educator Evaluation System (NCEES). The desired outcome was to make sense of teacher perspectives about teacher evaluation and how the process informs practice. Therefore, a phenomenological approach was the most fitting methodology. The study was conducted through semi-structured interviews with participants. Interviews usually lasted 45 minutes to an hour.

Participants

This study was conducted in a large, urban school district in North Carolina. Participants represented a variety of teacher demographics within that district. Specifically, participants varied by race, gender, years of experience, grades and subjects taught, and school setting. Participants also had a variety of life experiences. For example, some participants had previous careers, and some participants had previous experience working in other states. Thirteen of the participants had earned career status and 10 had not. Two of the participants were African-American females, 18 were

Caucasian females, and three were Caucasian males. Nine of the participants taught math, 16 of the participants taught reading, and 11 of the participants taught science. Four participants taught fourth grade, six participants taught fifth grade, three participants taught sixth grade, three participants taught seventh grade, two participants taught eighth grade, and five participants taught more than one grade between sixth and eighth grades.

Because this study's focus was on the shared experiences of teachers evaluated with the NCEES and because I wanted to include teachers from a variety of settings and demographics, I used purposeful sampling. I believed this method to be the most appropriate approach to participant selection since I hoped that the study might provide insight into the experiences, perceptions, and understanding of other teachers in North Carolina being evaluated using the NCEES using an inductive approach to data analysis.

Instrumentation

A guided or semi-structured interview was used to collect data. The semi-structured interview was chosen as the method to collect data because each individual's lived experience with the teacher evaluation process is unique and therefore some thoughts, feelings, or perceptions had to be clarified or explored beyond the interview protocol. The semi-structured interviews were conducted between April 3, 2014 and June 16, 2014. The interview protocol is available in Appendix B.

Research Procedures

I approached three elementary school principals, three middle school principals, and one principal in a K-8 school to obtain consent to recruit teachers for my study at their respective schools. I selected these schools so that a variety of grade levels and

school demographics would be represented. Each of the principals I approached about my study initially agreed to allow me to recruit teachers at his or her school. However, one of the middle school principals who consented to his school being a recruiting site did not respond to my attempts to schedule a recruiting session at his school. Therefore, I conducted recruitment sessions at three elementary schools, two middle schools and one K-8 school.

Table 2 provides a list of each school with the number of recruits consenting to being participants followed by the number of teachers who actually participated in the study. One recruit at Delp Magnet stated that she did not want to be interviewed until the Fall, well outside of the data collection window. Therefore, she did not participate. One recruit at Goudreau Elementary communicated via email that she had changed her mind and no longer wished to participate. A recruit at Hashian Elementary was not eligible to participate because she was not being evaluated on the NCEES. Other recruits who did not participate in the study, four recruits, simply did not respond to my attempts to schedule an interview, despite their initial interest in the study. A total of 23 teachers participated in the study out of an initial recruiting pool of 30.

Of the six participating schools, three schools were elementary serving students in Pre-Kindergarten (Pre-K) through fifth grade, two schools were middle schools serving students in sixth through eighth grade, and one school serving students in grades Kindergarten through eighth grade. Three of the six schools were traditional schools and three were magnet schools. Two of the six schools had Title I status.

Table 2

Recruiting Participants by School

School*	Description	Number of Initial Recruits	Number Who Actually Participated
Scholz Elementary	Elementary (PreK-5)	4	4
Delp Magnet	K-8	5	3
Sheehan Middle	Middle (6-8)	4	4
Masdea Middle	Middle (6-8)	7	7
Goudreau Elementary	Elementary (PreK-5)	5	3
Hashian Elementary	Elementary (PreK-5)	5	2

*Pseudonyms are used.

The composite scores of participating schools ranged from 62.6 to 99.8%. Composite scores refer to the percentage of test scores that are at or above a proficient level. Participating schools met 86.2% to 100% of their Annual Measurable Objectives (AMOs). AMOs are performance objectives that North Carolina has set for each school to meet requirements of *No Child Left Behind.* Objectives exist for proficiency of students of different racial backgrounds, economic status, limited English status, and disability status. Objectives also exist for attendance, participation, and proficiency of all students as a whole (www.ncpublicschools.org/accountability).

According to the Department of Public Instruction's Accountability Model, one of the schools was an Honor School of Excellence, two schools were designated as School of Distinction, two schools were designated as School of Progress, and one school was identified with No Distinction. Four schools made high growth, one school met expected

growth, and one school did not meet expected growth. Table 3 provides a description of each recruiting site.

Table 3

Site Demographics

School	Grades	Perf. Comp.	AMO Status	ABC Status	Title I	Percentage of Free/ Reduced Lunch Population*	Type of School
Scholz	PreK-5	82.4%	100% of AMOs met	School of Distinction; High Growth	No	40%	Traditional
Delp	K-8	75.2%	87.5% of AMOs met	School of Progress; High Growth	Yes	77%	Magnet
Sheehan	6-8	99.8%	100% of AMOs met	Honor School of Excellence; Met Expected Growth	No	16%	Magnet
Masdea	6-8	80.7%	82.8% of AMOs met	School of Distinction; High Growth	No	43%	Traditional
Goudreau	PreK-5	78.3%	100% of AMOs met	School of Progress; High Growth	No	52%	Traditional
Hashian	PreK-5	62.6%	86.2% of AMOs met	No Recognition; Did Not Meet Expected Growth	Yes	88%	Magnet

*This information was obtained from www.ncpublicschools.org/fbs/resources/data/

I recruited participants by visiting each school site at a time and date arranged with the principal of each school. Eligible participants within the school were invited to hear a 10-minute presentation of a description of my study. During this presentation, I

used the oral recruitment script, which appears in Appendix A. Interested participants put their name, email address, and phone number on an index card and placed them on a table for me to collect after participants left. Uninterested participants were instructed to leave their index card blank.

Within a week of each recruiting session, I emailed each participant who provided contact information. Participants responded with a date and time of their convenience for us to meet to conduct the interview. Two of the 23 participants chose to meet at locations other than their work site. All other participants chose to meet at their work site after school hours. Interviews were audio recorded and labeled with a number that represented the chronological order by which the interview took place. For example, the ninth interview was labeled, "Transcription #9."

Interviews typically lasted approximately 45 minutes to one hour in length. Cumulatively, the interviews were more than 16 hours of dialogue time. The shortest interview lasted only 22 minutes and 20 seconds. The longest interview lasted 71 minutes and 54 seconds. Participants who had little knowledge or recollection about standards on the NCEES tended to respond to that segment of the interview with short answers indicating that they could not recall elements of the standard. Therefore, those interviews tended to be shorter in length.

For each of the interviews, I used a MacBook Pro to record the interviews using QuickTime Player. I set up the computer so that participants could see the computer screen and see that they were not being video recorded. I had to pause recording on three of the interviews due to interruptions. In hindsight, the interviews that occurred off

campus were quite short and I wonder if the settings were too busy for participants to feel comfortable elaborating on their responses.

I began each session by engaging each participant in general conversation. I asked them about their day, since all but one interview occurred after work during the week. This usually began an open dialogue that we shared for five to ten minutes before we reviewed the consent form. I allowed each participant to read the consent form while I retrieved my computer and interview protocol from my bag and prepared for the interview. I asked participants if they had any questions about the consent form and wanted any additional information before they signed.

During the interviews, several participants displayed some verbal and/or nonverbal signs of embarrassment when they could not remember the standards on the NCEES or elements of the various standards. Therefore, at times I provided verbal reassurance that it was typical for participants to not recall elements within the standards or even the standards themselves. This reassurance seemed to alleviate participants' discomfort. The first three participants I interviewed were teachers I had worked with previously during prior academic years and, therefore, they were more vocal in sharing with me that they felt embarrassed for not being able to recall the standards. For example, the first participant stated that she was glad she knew me because she would have felt "really dumb" in front of a stranger. Due to this feedback, I feel that I was more sensitive to subsequent participants' feelings and was more proactive in telling them in advance that it was not unusual for participants to not recall details about the evaluation instrument.

Most of the participants had negative feedback about the evaluation process, the evaluation tool, quality of feedback, current or previous evaluators, or the general climate of teaching in North Carolina. Therefore, I believe that most participants felt comfortable enough to share their honest opinions and feelings about teacher evaluation and about the NCEES. Also, once the interviews were over, most participants stated that they were happy to help with my study or even encouraged me to contact them if I needed anything else from them.

After interviews were completed, I had them transcribed using a professional transcriber. Once the transcriptions were completed, I conducted member checking by contacting participants via email to provide them with the opportunity to review their interview transcription. Four participants asked to review their interview transcripts. Three asked to have the content emailed to them. One participant asked to have it mailed to them. One of the four participants wanted some information stricken from her transcription that she felt could possibly identify her. I deleted the requested information from her transcription. The other three participants who wanted to view their transcription did not have any concerns. The remaining 19 participants did not reply to my invitation to view their transcription, but one of those participants asked to read "the final analysis." I contacted that participant to clarify exactly what she wanted and she stated she was just interested in the final product and to know what teachers thought about the topic. Therefore, she did not want to check the data for review purposes, only for her own personal interest in the subject matter.

Once participants had the opportunity to review and finalize their input, I created analytic tables, recording each participant's responses to each of the interview questions. This allowed me to codify responses easily. On the analytic tables, I coded key phrases that I then placed into categories. The analytic tables also allowed me to view participant responses as I considered participant demographics, such as years of experience.

In addition to soliciting approval from participants on their own interview transcription, I also selected 2 professional colleagues to review the interview transcriptions and the analysis of the data to check for interviewer bias during the interview process. These two professional colleagues were African-American females. One of the peers was a school administrator and one of the peers was a teacher working in a leadership role within the school district where the study took place. I obtained a statement of confidentiality from each of these colleagues.

The peer review session lasted an hour. Each peer reviewer was given a one-page executive summary of the NCEES (Table 1). They also received a summary of school demographics (Table 2) and copies of the Oral Recruitment Script (Appendix A). I explained to the peer reviewers the purpose of the study and the details of the recruitment process. Then, I explained the criteria for participant selection and the demographics of individuals who agreed to participate in the study. They also received a summary of participant demographics (Appendix D). Next, we viewed the interview protocol (Appendix C). I explained to peer reviewers how interviews were transcribed and key statements transferred to analytical tables. Peer reviewers had the opportunity to view

transcriptions and corresponding analytical tables. Finally, I explained how I was using the analytical tables to identify categories and themes in the data.

Peer reviewers reported that the recruitment process seemed sound. One peer reviewer wanted more information on how schools were selected. I explained to her that I researched school demographics and selected an initial group of schools that represented a variety of these demographics. My plan was to select additional schools if I could not get enough recruits from the initial group of schools. Peer reviewers gave positive feedback on the analytical tables as a tool for organizing key statements from transcriptions. In terms of analysis of the data, they viewed the data tables, included in this publication, and were able to see how the information was transferred from the transcription to the analytical table to the data table. One peer reviewer asked if the final publication would include a comparison of participants' perceptions of fairness from Title One versus non-Title One schools and I explained that the final publication would include an examination of all school and participant characteristics related to any existing trends in participant responses.

Data Analysis

Once interviews were transcribed, I created an analytical table for each interview question to record key information contained within participants' responses. From these analytical tables, I codified participant responses and categorized them by content. Then, I identified common themes across participant responses. Lichtman (2010) describes this iterative process of data analysis as a common method in qualitative research.

For some of the participant responses, I categorized them by level of performance according to the NCEES. I used specific descriptors of each standard, element, and performance level to classify responses in this manner. I explained why I classified each response as I did in an attempt to be transparent about my analysis of participants' perceptions of the standards on the NCEES.

Assumptions of the Study

It is assumed that the participants in this study reflect a range of teacher characteristics, school settings and demographics, and experiences so that some insight may be gained about the common, lived experiences among teachers with the teacher evaluation process and with the use of the NCEES in North Carolina. It is also assumed that the participants' understanding of and perceptions about each standard of the NCEES found in this study may be indicative of the general understanding of and perceptions about the NCEES standards among teachers in North Carolina. Furthermore, it is assumed that the quality of feedback received by participants from their evaluators may also be indicative of the quality of feedback received by teachers in North Carolina and that the impact on participants' professional practices may also be indicative of the impact of feedback on teachers' professional practices in North Carolina.

Trustworthiness

To address the issue of trustworthiness, each participant was provided with a consent form that explained the purpose and breadth of the study. All interviews were audio-recorded and all recordings were saved so that transcriptions could be checked for accuracy, if requested. After transcriptions were completed, I gave all of the participants

the opportunity to review the data and to provide feedback or any additional input. This practice is referred to as member checking (Lichtman, 2010). All data was kept securely and participants' anonymity was maintained. In addition, I utilized three peer reviewers who critiqued my methodology, analytical tables, and findings. I attempted to be transparent in my conclusions by explaining my thinking when classifying participant responses according to the NCEES and by providing specific examples of participant responses and thick descriptions throughout the analysis of data from participant interviews.

Statement of Subjectivity

As a school administrator, I have worked with many teachers with varying levels of competency and effectiveness. In 1995, I completed a graduate program, earning a Master of Psychology and an Educational Specialist in School Psychology degree. In that program, objectivity and quantitative data were emphasized. Likewise, qualitative research was de-emphasized. Therefore, my philosophy has been grounded for a long time with that which is objective and measureable. I have supported the concept of teacher accountability using student growth data, as well as a thorough method of teacher evaluation. However, the results of the study, provided in detail in Chapter IV, suggest that neither of these practices in North Carolina are influencing teacher practices or promoting significant reflection. Therefore, my subjectivity has not influenced the participants or the outcomes of the study. Rather, my participants and the study have influenced my subjectivity.

Limitations of the Study

Participants were chosen from elementary and middle schools within one district in North Carolina. Therefore, results of the study cannot necessarily provide insight into the perspectives of teachers from other grade levels or from other states. Since all districts in North Carolina use the NCEES, findings from this study may be relevant to the experiences of teachers within other districts of North Carolina. However, evaluator training, district philosophies, or other factors may influence teachers' experiences and perceptions related to the teacher evaluation process and the NCEES in ways that are different from the participants in this study.

Summary

This chapter described the research design and methodology of the study. The research design is a qualitative design and the methodology used is a phenomenological approach. This chapter provided the rationale for this design and methodology. The recruitment of participants from a variety of schools was also outlined and demographics of recruiting sites were provided. The use of a semi-structured interview was explained and rationalized. Procedural details of the study were specified, including how interviews were conducted and transcribed. Participants reviewed transcriptions and content of transcriptions were transferred to analytical tables, which were then used to analyze the data. The peer review process was explained as well. It was assumed in this chapter that results from this study may provide insight into the perceptions and experiences of teachers in North Carolina related to the teacher evaluation process and the use of the NCEES, though limitations of the study were acknowledged. Finally, the chapter

provided an explanation on the attempt to ensure trustworthiness of the study and a statement of subjectivity was also provided.

The next chapter, Chapter IV, provides a complete discussion of the study's findings. Specifically, Chapter IV will reiterate the rationale for the study. The study's participants will be described in more detail. Results from the collection of data will be discussed as they relate to the research questions. The next chapter will provide information about the knowledge that participants have about the standards on the NCEES, their perceptions on the fairness of the instrument, how they have been rated on the NCEES, what type of feedback they have received, and the extent to which the feedback has impacted professional practices. Their perceptions on the fairness of the teacher evaluation process in general will also be provided. In addition, participants' knowledge of and perceptions about VAD and its inclusion in the NCEES will also be shared. Finally, Chapter IV will provide a description of recommendations that participants offer for improving the teacher evaluation process in North Carolina and alternative ways to evaluate teacher effectiveness that participants suggest.

CHAPTER IV

RESULTS

The rationale for the study was to provide insight into the new evaluation tool in North Carolina, the NCEES, and its impact on teacher practices in the classroom. Teacher perceptions and understanding of the evaluation tool and of the evaluation process will directly affect the extent to which they will use the standards-based feedback and the value-added data to reflect upon their professional practices and make changes for potential improvements. In addition, understanding how teachers perceive the evaluation process and the evaluation tool might provide insight into what improvements may be necessary to increase the efficacy of teacher evaluation in North Carolina and how concerns may be addressed as we move forward with policy decisions that affect practices related to teacher accountability.

This study sought to answer the question: What are teachers' experiences and perceptions about the evaluation process using the NCEES and how do these experiences and perceptions inform their practices? More specifically, this study sought to find out what feedback teachers receive on the NCEES and how it impacts their practices, what teachers understand about the standards on the NCEES, what teachers think about the fairness of the standards-based portion of the NCEES, and what teachers think about the fairness of VAD being included in the NCEES. Finally, this study sought to find what recommendations teachers might offer about the standards, the instrument, or the

evaluation process in general and what suggestions they might have about ways to evaluate teachers and hold them accountable for student achievement.

The overarching research question was: What are teachers' experiences with and perceptions about the evaluation process using the NCEES and how do these experiences and perceptions affect the extent to which the evaluation process informs their practices? The questions undergirding this overarching question were: 1) What do teachers understand about Standards One through Six of the NCEES? 2) What do teachers think about the fairness of the standards-based portion of the NCEES? 3) What do teachers think about the fairness of value-added data as a component of the NCEES? 4) What recommendations do teachers offer about each standard, the NCEES, and methods of teacher accountability for student performance?

Twenty-three participants from six schools were successfully recruited to participate in this study. Participants' years of experience in teaching ranged from 1 to 22 years. Sixty-one percent of the participants had tenure status and 39% of the participants did not. Fifty-six percent of the participants taught at the middle school level and 44% of the participants taught at the elementary school level. Sixty-one percent of the participants had advanced degrees or were working on advanced degrees. Thirty percent of the participants had worked in another profession other than teaching. Twenty-two percent of the participants previously taught in a state other than North Carolina.

All participants taught at least one subject for which value-added data is derived. The majority of participants, 70%, taught reading. Forty-four percent of the participants taught math, 44% taught science, and 39% taught social studies. Fifty-seven percent of

participants taught more than one subject. Seventeen percent of the participants taught fourth grade and 26% of the participants taught fifth grade. Therefore, 43% of the study's participants taught at the elementary school level. Thirteen percent of participants taught sixth grade, 9% taught seventh grade, and 9% taught eighth grade. Twenty-two percent of participants taught more than one grade level at the middle school level. Overall, 57% of the study's participants taught at the middle school level. Appendix D provides a summary of participant demographics.

A total of 33 questions comprised the interview protocol. The first three questions were general questions, asking participants about their level of experience, tenure (career) status, and educational background. I also asked each participant to identify grades and subjects in which they were currently teaching. The remaining interview questions were categorized into four areas: knowledge of standards-based teacher evaluation, knowledge of value-added measures, impact on teacher practice, and improving the teacher evaluation process. These four areas corresponded with the research questions.

General Perceptions and Experiences of Participants

Table 4 summarizes participant responses on interview questions relating to general perceptions of and experiences with the evaluation process and with the NCEES. Twenty-two categories emerged from participant responses. The category with the most responses was the category, "Just a Requirement / Not Important." Forty-eight percent of participants made statements about the evaluation process as being unimportant in some way or little more than a requirement that needs to be completed, void of meaning or application.

Table 4

General Perceptions and Experiences of Participants by Category

	Accountability is Necessary	**Positive Experiences**	**Fair**	**Just a Requirement/Not Important**	**Subjective**	**Snapshot**	**Limited Feedback**	**VAD Affects Observation Ratings**	**VAD is Unfair/Problematic**	**Too Time Consuming for Administrators**	**Too Lengthy**
P1	X		X	X	X						
P2	X					X					
P3			X								X
P4					X	X					X
P5				X			X				X
P6			X				X				X
P7				X	X						
P8											
P9					X		X				X
P10						X				X	
P11	X			X						X	
P12											
P13				X							
P14	X			X					X		
P15	X					X		X	X		
P16		X									
P17											
P18				X	X		X			X	
P19	X			X	X					X	
P20				X							
P21			X	X			X				
P22				X	X		X				
P23			X						X		
Totals	**6**	**1**	**5**	**11**	**7**	**4**	**6**	**1**	**3**	**4**	**5**

Table 4

(Cont.)

	Low Morale/ Apprehension	Components Not Observable	Vague	Teachers Don't Understand It	Unrealistic	Growing Tool	Need Multiple Formats	Clarity of Expectations	Erroneous Statement	Specific	Being Misused
P1					X		X				
P2	X	X									
P3					X		X				
P4			X	X	X		X				
P5					X						
P6								X			
P7	X	X				X					
P8	X	X					X				
P9		X			X						
P10					X			X	X		
P11											
P12	X						X				
P13											
P14	X										
P15					X						
P16											
P17						X		X		X	
P18					X						X
P19					X						
P20		X								X	
P21								X		X	
P22											
P23		X									
Totals	**5**	**6**	**1**	**1**	**9**	**2**	**5**	**4**	**1**	**3**	**1**

For example, Participant 5 stated that she is "not really caring what anybody else is saying" about her practices. Throughout the interviews, participants referenced a lack of feedback from the evaluation process, particularly when directly questioned about feedback. However, several participants identified this as a problem with the process

when asked to share their general perceptions in the very beginning of the interview. Participant 18 stated, "It's kind of something that has to happen, rather than something that will help [teachers] become stronger." Participant 22 clarified that she is frustrated with the lack of feedback by saying, "I'm insulted when I get proficient and I'm not hearing what I could do . . . to get better." This reference to a general lack of useful feedback is prevalent and will be discussed further in this chapter.

In addition to these comments, additional participants described the evaluation process as "just one more thing that we don't spend enough time on" (Participant 11) and something that needs to be taken "with a grain of salt" (Participant 14). Participant 22 said, "I don't see how it helps with anything" and Participant 13 said that the evaluation process is "more like a deadline. We have to get this done by this deadline." Many of the participants who expressed a lack of importance on the evaluation process stated that it was primarily due to the subjective nature of standards-based evaluations. Participant 4 argued, "You can't get away from the subjectivity." Participant 9 stated that how you are rated "really depends upon who your principal is." One of Participant 18's responses elaborates upon that perception. She stated that "if [teachers] don't have a strong relationship with the evaluator, then they think it's biased and that it's very subjective." Participant 22 said, "It's . . . opinions about people, subjective."

Another frequent response shared by participants when asked about their general perceptions of the evaluation too was a perception of the tool being too lengthy and/or too time consuming. For example, Participant 5 stated that the instrument was "too lengthy" and "just difficult to navigate through all of the parts and pieces." Another

participant (#9) described it as "cumbersome." Participant 18 noted that it is so time consuming, that evaluators do not do what they are supposed to do. She stated that she was supposed to be observed three times during the year but was only observed two times and had not been given the opportunity to have a pre-observation conference or a post-observation conference all year.

Many of the participants felt that the NCEES is unrealistic in all that it expects from teachers. As Participant 8 explained, "It's difficult to master everything." Participant 9 expressed significant frustration with the instrument when she said, "The number of areas that we're evaluated in is absurd." Another participant (#11) said that she understood that the "purpose is to make sure teachers are engaging in best practices, but we're so overwhelmed . . ." Other responses that had to do with the length of the instrument and unrealistic expectations of the instrument had to do with the participants' opinions that too many items on the instrument are not observable during a typical classroom walk-through. Some participants, in turn, said this contributes to a low morale among teachers.

Knowledge of Standards-Based Evaluation

Standard One: Teachers Demonstrate Leadership

Surprisingly, participants had a difficult time identifying the standards of the NCEES. Eighty-three percent of participants could not identify Standard One of the NCEES. Even after I told participants that Standard One was Teachers Demonstrate Leadership, the majority of participants could only identify one or two elements of leadership. Most participants (87%) were able to identify components of leadership

within the school or, at least, identify that leadership within the school was an element of Standard One. Thirty-five percent of participants identified components of leadership within the classroom. However, only 13% of participants identified components of leadership within the teaching profession and none of the participants identified the elements of leadership related to advocating for schools and students or related to demonstrating high ethical standards. Table 5 summarizes these findings.

Of the descriptions that participants provided about Standard One, Element A: Teachers lead in their classrooms, only one of the participants identified a specific descriptor under Element A. Rather, participants tended to state the element itself. Participant 8 stated that a teacher demonstrating appropriate classroom management would be an example of Standard One. This example would fall under the Proficient level of Standard One, Element A because it describes a safe and orderly classroom, but does not include a component of student collaboration.

Of the descriptions that participants provided about Standard One, Element B: Teachers demonstrate leadership in the school, three participants stated the element itself. Of the remaining descriptions, one of them, provided by Participant 9, fell in the Developing category ("going to meetings"). This description indicates a Developing level of performance because it refers to attending meetings, but does not include active participation in meetings. Ten descriptions fell in the Proficient category, some of which were stated by more than one participant. These were: "being on committees," "speaking out in meetings," "taking part in committees," "working with colleagues," "participating in meetings," "serving on committees," "helping develop the school improvement plan"

and "voicing your opinions." These descriptors fall within a Proficient level of performance because they all refer to active participation, but do not refer to assuming a leadership role.

Table 5

Knowledge of Standards-Based Evaluation and Perceptions about Standards, Standard One

	Classroom	**School**	**Profession**	**Advocate**	**Ethics**	**Unrealistic**	**Need Multiple Formats**	**Components Not Observable**
P1						X	X	
P2	X	X	X					
P3						X	X	
P4						X	X	
P5		X						
P6		X						
P7		X						X
P8	X	X						
P9		X						
P10		X						
P11	X	X						
P12		X						
P13		X						
P14	X	X						
P15	X	X	X			X		
P16	X	X	X					
P17		X						
P18		X				X		
P19		X						
P20	X	X						X
P21	X	X						
P22		X						
P23		X						
Totals	**8**	**20**	**3**	**0**	**0**	**5**	**3**	**2**

Twelve descriptions fell in the Accomplished category, some of which were stated by more than one participant. These were: “facilitating meetings,” “promoting the school improvement plan,” “helping other teachers,” “being a leader with colleagues,” and “leading committees.” Each of these descriptors include a leadership quality in the teacher’s participation. Two descriptions provided by participants fell in the Distinguished category. These were: “encouraging colleagues to change their practices and promoting the school improvement plan” and “mentoring others and teaching your colleagues.” Both of these descriptors describe a teacher’s impact on others’ practices within the school.

Of the descriptions that participants provided under Standard One, Element C: Teachers lead the teaching profession, all three referred to leadership outside of the school. These were: “present things on a county or state level,” “leadership outside of the school,” and “leadership for the district.” These fell in the Distinguished category because they all refer to leading professional development or decision-making beyond the school level.

Overall, therefore, five participants (22%) were able to provide a description of Standard One that is at a Distinguished level of performance. Twelve participants (52%) were able to provide a description of Standard One that is at an Accomplished level of performance. Eleven participants (48%) were able to provide a description of Standard One that is at a Proficient level of performance. One participant (4%) provided a description of Standard One that is at a Developing level of performance. Three participants (13%) did not provide any descriptions of Standard One. See Appendix E for

a summary of participant perceptions of Standard One as they relate to levels of performance.

Twenty-two percent of participants perceived Standard One as unrealistic. Three participants specifically referred to this standard as being unfair to newer teachers, who they feel do not typically have as many opportunities to have leadership roles outside of the classroom. Participant 1 said, "I'm a beginning teacher, so I can't even really live up to that standard." Participant 3 said, "It's harder for a newer teacher to show some of that leadership . . . You don't really have the opportunity as a brand new teacher." Participants 7 and 20 noted that leadership is not always observable to evaluators. For example, Participant 7 noted, "A principal may never see me collaborate with a bunch of teachers, helping them." Participant 15 said, "There's leadership within the classroom and there's leadership within the school community of peers as well as outside of the school. There's value in that, but we're not all leaders. Not everyone fits that mold."

Standard Two: Teachers Establish a Respectful Environment for a Diverse Population of Students

Table 6 depicts participants' knowledge of and perceptions about Standard Two. Only one participant out of 23 was able to identify Standard Two. After being reminded that Standard Two was about establishing a respectful environment for a diverse population of students, 22% of the participants could identify two of the five elements of performance under Standard Two. Sixty-one percent could identify one element of performance under Standard Two. Forty-eight percent of participants identified the element of providing a positive and nurturing environment for students and 48% of participants identified the element about embracing diversity in the classroom and in the

world. Two participants (8%) identified that working collaboratively with families was part of Standard Two. Only one participant mentioned adapting teaching for students with special needs and only one participant mentioned practices that would fall under treating students as individuals.

Of the descriptions provided by participants of Standard Two, Element A: Teachers provide an environment in which each child has a positive, nurturing relationship with caring adults, eight of them fell in the Proficient category. These were: "all children to be treated with respect," "treating students fairly," "to create an environment that makes every kid feel welcome," "to be respectful of everybody," "how I perform as far as my relationship with students and how accepting I am," "being fair to all students," "making your classroom a place where people feel respected and they feel comfortable with the learning environment," "building relationships with your students," and making an "environment they can work in, be accepted in, and respected." These descriptors fell in the Proficient category because they refer to establishing a fair or respectful relationship with or environment for students.

Three descriptions fell within the Accomplished category of Element A. These were: "maintaining a nurturing classroom," "making your classroom a place where people feel respected and they feel comfortable with the learning environment," and providing a "classroom climate of respect and a positive atmosphere in the class [where] everyone's accepted and included." These described an Accomplished level of performance because they all indicate the teacher's ability to maintain a positive classroom climate over time. None of the descriptions fell under the Developing or

Distinguished categories of Element A. Appendix F provides a summary of all descriptions and their corresponding level of performance on the NCEES.

Table 6

Knowledge of Standards-Based Evaluation and Perceptions about Standards, Standard Two

	Positive Environment	**Embrace Diversity**	**Students as Individuals**	**Adapt Teaching for Students with Special Needs**	**Work Collaboratively with Families**	**Easiest Standard to Master**	**Gives Incorrect Information**
P1				X			
P2	X	X					
P3	X	X					
P4	X		X		X		
P5	X						
P6						X	
P7							X
P8		X					
P9		X					
P10		X					
P11	X						X
P12						X	
P13	X	X					
P14	X						
P15		X					
P16		X					X
P17	X						
P18	X				X		X
P19		X					
P20		X					X
P21	X						
P22	X	X					
P23							X
Totals	**11**	**11**	**1**	**1**	**2**	**3**	**6**

Eleven descriptions from participants fell under Standard Two, Element B: Teachers embrace diversity in the school community and in the world. One of these fell in the Developing category. It was: "putting flags from every country in your room and . . . readings from various cultures." This descriptor falls in the Developing level of performance because it suggests an awareness of diversity, albeit a superficial one. Seven descriptions from participants fell in the Proficient category. They were: "respecting different religions and different cultures," "showing an understanding of diversity," "fairly encounter each student [regardless of] a learning difference, racial difference, etc. and that I meet students where they are culturally," "making sure the environment of my classroom welcomes diversity," "how well I recognize my students' diversity and how well I incorporate [their diversity] into my lesson plans," "that no matter what size a child is, what color they are, what sex they are, what their thinking is, what their abilities are, what kind of family they come from, you treat them equally, but equally doesn't mean the same because what one child needs may not be what another child needs," and "to make sure that everybody, no matter what kind of background or color or culture, feels accepted and included." These descriptors fall in the Proficient level of performance because they all refer to an active use of their knowledge of diversity.

Three descriptions provided by the participants fell in the Accomplished category. They were: "that you're respectful of different sources, especially with lesson plans and respectful of different viewpoints on topics," "embracing other cultures and diverse students," and "you give the students what they need individually, no matter where they come from culturally or economically. You're culturally sensitive, provide culturally

relevant material." These descriptions fall in the Accomplished level of performance because they each imply an appreciation of diversity that will be taught and promoted among the students. None of the participants provided descriptions of Standard Two that fell in the Distinguished category of Element B.

Only one description of Standard Two fell under Element C: Teachers treat students as individuals. It fell in the Developing category and was: "holding high expectations." This was at a Developing level of performance because it acknowledges the importance of high expectations but does not refer to the communication of expectations to students. (At the developing level, the NCEES states, "Holds high expectations of students.")

Additionally, only one description of Standard Two fell under Element D: Teachers adapt their teaching for the benefit of students with special needs. It fell in the Developing category because the statement acknowledges the importance of knowing the learning needs of the students and includes a basic response to different needs. It was: "understanding your students' learning styles, modifying their work."

Two participants provided descriptions of Standard Two that fell under Element E: Teachers work collaboratively with the families and significant adults in the lives of their students. Both descriptions fell in the Proficient category and were identically stated by each participant: "communicating with families." This response indicates a Proficient level of performance because it indicates an initiation of communication but does not include the awareness of and need to overcome obstacles to communicating with families.

Overall, none of the participants were able to provide a description of Standard Two at the Distinguished level of performance. Five participants (22%) were able to provide one descriptor that fell in the Accomplished level of performance. Thirteen participants (57%) were able to provide one or two descriptors that fell in the Proficient level of performance. One participant (4%) provided a descriptor that fell in the Developing level of performance. See Appendix F for a summary of participant perceptions of Standard Two as they relate to levels of performance.

Seven participants (30%) provided descriptions that did not fall under Standard Two, suggesting that they have an overall lack of understanding of the standard. For example, Participant 7 stated that a practice demonstrating Standard Two is "that a teacher can run a classroom." Participant 23 said that classroom management and organization falls under Standard Two. Another participant (#16) felt that 21st century skills and dispositions fall under Standard Two. Participant 18 actually provided an example that falls under Standard One, which was making sure that students collaborate and lead in the classroom. Participant 20 erroneously identified that an example of a teaching skill that falls under Standard Two would be making sure that the students are respectful of the teacher.

Standard Three: Teachers Know the Content They Teach

When asked to identify Standard Three of the NCEES, 70% of participants could not recall. Twenty-six percent of participants asked if the standard had something to do with content. Only one participant stated confidently that Standard Three was about knowing the content. Once participants were reminded of the standard, 65% of

participants provided examples of practices that fall under the element of content knowledge. Thirty percent provided examples of aligning instruction with curriculum standards. However, only 17% provided an example of recognizing the interconnectedness of content areas and only 8% provided an example of making instruction relevant to students. In addition, only 35% of participants identified more than one element under Standard Three. The majority only identified one of the four elements under Standard Three. Therefore, none of the participants seemed to have a full understanding of Standard Three. Table 7 summarizes these findings.

Of the descriptors provided by participants that fell under Standard Three, Element A: Teachers align their instruction with the North Carolina Standard Course of Study, most of them fell in the Developing category. They were all stated as either "being aware of the standard course of study" or "knowing the standard course of study." These are examples of a Developing level of performance because they simply refer to an awareness of the standards. Two of the descriptors fell in the Proficient category. They were: "teaching based on the standard course of study and common core" and "knowing the curriculum for your teaching and having all the knowledge you need to teach it." These indicate a Proficient level of performance because both descriptions refer to the utilization of the standards in planning and teaching. None of the descriptors fell in the Accomplished or Distinguished categories.

Table 7

Knowledge of Standards-Based Evaluation and Perceptions about Standards, Standard Three

Standard Three: Teachers Know the Content They Teach							
	Aligned Instruction	**Content Knowledge**	**Content Interconnectedness**	**Makes Instruction Relevant**	**Easy to Master**	**Most Important Standard**	**Restricted by School District**
P1		X					
P2	X						
P3				X			
P4	X		X				
P5	X		X				
P6		X	X			X	
P7		X					
P8	X	X					
P9		X	X				
P10		X					
P11		X					
P12					X		
P13		X					
P14	X	X					
P15		X				X	
P16		X					
P17		X		X			
P18	X						
P19							X
P20	X	X			X		
P21					X		
P22		X					
P23		X			X		
Totals	**7**	**15**	**4**	**2**	**4**	**2**	**1**

Of the 15 participants who provided a description of Standard Three, Element B: Teachers know the content appropriate to their teaching specialty, 14 of them just stated

the element of knowing the content. One participant provided a response that falls in the Proficient category. It was: being "competent in the subject matter that I teach and that I'm qualified and that when I stand up and teach I know what I'm talking about and I know how to present the information." This indicates a Proficient level of performance because it refers to the teacher's knowledge of content and the utilization of that content knowledge in teaching practices.

Four participants provided a description of Standard Three that fell under Element C: Teachers recognize the interconnectedness of content areas/disciplines. Two of the descriptors fell in the Proficient category. They were: "vertical planning" and "when you're teaching a subject, you understand where the kid came from and where they are going next—you understand that curriculum." These descriptors fall within the Proficient level of performance because they refer to utilizing vertical awareness in planning and teaching. Two of the descriptors provided by participants fell in the Accomplished category. They were: "I know the actual content of what I'm teaching and I can link it to prior grade levels as well as link it to other content" and "knowing your content and being able to make connections to and integrating other subjects." These indicators fall in the Accomplished level of performance because they both refer to linking content to other areas of content or disciplines.

Two participants provided a description of Standard Three that fell under Element D: Teachers make instruction relevant to students. These descriptions fell in the Proficient category. They were: "incorporating 21st century" and "it's more about, you know your content but also how to help learners become 21st century learners." These

are proficient examples because they indicate more than just an awareness of 21st century content; they imply using that knowledge during instruction.

Overall, none of the participants provided a description of Standard Three that fell at the Distinguished level of performance. Two participants (8%) provided a descriptor of Standard Three that fell at the Accomplished level of performance. Six participants (26%) provided a descriptor of Standard Three that fell at the Proficient level of performance. Five participants (22%) provided descriptors that fell at the Developing level of performance. See Appendix G for a summary of participant perceptions of Standard Three as they relate to levels of performance.

Additionally, a few participants shared a few extraneous thoughts about Standard Three. Four participants (17%) felt that Standard Three was easy for teachers to master. Two participants (8%) identified Standard Three as the most important of the standards. One participant (4%) conveyed that teachers are too restricted by the district in terms of content.

Standard Four: Teachers Facilitate Learning for Their Students

Standard Four is the lengthiest standard of the NCEES. None of the participants in the study could identify Standard Four. Once they were told what Standard Four is, 26% of participants identified differentiation as being part of Standard Four. Eight percent provided examples that fell under the element of planning appropriate instruction for students. Seventeen percent of participants provided examples of using a variety of instructional methods. Eight percent mentioned the use of technology, 8% mentioned students working in teams, 4% or one participant mentioned effective communication

skills and one participant mentioned developing critical-thinking skills in students. Twenty-six percent offered perceptions of Standard Four that related to using a variety of assessment methods. One participant described four of the eight elements of Standard Four. Seventeen percent of participants described two of the eight elements. Fifty-seven percent of participants could only describe one element of Standard Four. Twenty-two percent of participants could not describe any of the elements of performance under Standard Four, even after being told the standard. Table 8 summarizes these findings.

Table 8

Knowledge of Standards-Based Evaluation and Perceptions about Standards, Standard Four

Standard Four. Teachers Facilitate Learning for Their Students								
	Knowledge of Learning Process and Developmental Levels	**Plans Appropriate Instruction**	**Uses a Variety of Instructional Methods**	**Utilizes Technology**	**Develops Critical Thinking and Problem-Solving Skills**	**Students Work in Teams**	**Communicates Effectively**	**Uses a Variety of Assessment Methods**
P1			X					
P2	X							
P3	X							
P4	X				X	X		X
P5						X		X
P6			X					
P7								
P8								X
P9	X							

Table 8

(Cont.)

Standard Four. Teachers Facilitate Learning for Their Students								
	Knowledge of Learning Process and Developmental Levels	**Plans Appropriate Instruction**	**Uses a Variety of Instructional Methods**	**Utilizes Technology**	**Develops Critical Thinking and Problem-Solving Skills**	**Students Work in Teams**	**Communicates Effectively**	**Uses a Variety of Assessment Methods**
P10								X
P11		X	X					
P12		X						X
P13								
P14				X				
P15								
P16								
P17	X							
P18				X				
P19								
P20	X							
P21			X					
P22							X	
P23								X
Totals	**6**	**2**	**4**	**2**	**1**	**2**	**1**	**6**

Of the participants who provided a description of Standard Four that fell under Element A: Teachers know the ways in which learning takes place, and they know the appropriate levels of intellectual, physical, social, and emotional development, half of them provided a descriptor that fell in the Developing category. These were: “I guess that’s where the differentiation comes in,” “a teacher is aware of differences in her students,” and “you can’t meet all these standards and I think this is one of them—

individualized learning." These indicate a Developing level of performance because they indicate an awareness of the practice. Two participants provided a descriptor that fell in the Proficient category. Those were: "doing differentiated lessons and meeting the needs of individual students" and "differentiating, meeting them where they are." These descriptors are at a Proficient level of performance because they refer to actively differentiating instruction for students. One participant provided a descriptor of Standard Four that fell in the Accomplished category of Element A. It was: "differentiating for your students who are lower, for your students who are higher, so that each student can access the curriculum and be successful with it." This descriptor is at an Accomplished level of performance because it refers to differentiation across the curriculum. None of the participants provided a description that fell in the Distinguished category of Element A.

Two of the participants provided perceptions or descriptions of Standard Four that fell under Element B: Teachers plan instruction appropriate for their students. Both descriptors fell in the Proficient category. These were: "using data to reflect" and "using data to make sure we're teaching appropriately." These descriptors were at a Proficient level of performance because they refer to the use of data to inform planning or teaching.

Four participants provided descriptions of Standard Four that fell under Element C: Teachers use a variety of instructional methods. Three of these descriptors fell in the Proficient category. They were: "teachers are not doing old practices, lecturing or just sitting there," "teaching in different ways, focus on how students learn best, be able to adjust," and "ability to present material to the kids in a way they understand it." These

descriptions of performance fall in the Proficient category because they include the utilization of different teaching practices. One participant provided a description that fell in the Accomplished category. It was: “when [students] cannot do it, we go back and reteach it in a variety of ways and [provide] options so your advanced learners can go further with the skill.” This description falls within the Accomplished level of performance because it describes the teacher ensuring the success of all students by using a variety of methods.

Two participants provided descriptors of Standard Four that fell under Element D: Teachers integrate and utilize technology in their instruction. Both of these did not fall under a specific level of performance; they were both statements posed as questions. They were, “Technology?” and “Does this go into technology?”

One participant provided a perception of Standard Four that fell under Element E: Teachers help students develop critical-thinking and problem-solving skills. It fell in the Proficient category. It was: “giving children the opportunity for feedback and helping them be critical thinkers.” This descriptor falls in at a Proficient level of performance because it describes a teacher actively promoting critical thinking without describing any specific types of critical thinking.

Two participants provided descriptions of Standard Four that fell under Element F: Teachers help students work in teams and develop leadership qualities. These descriptions fell in the Accomplished category. They were: “using student-led teams” and “doing things that build student learning communities, giving some of the control back to

them." This falls within an Accomplished level of performance because they both refer to students managing their own learning teams.

One participant provided a descriptor of Standard Four that fell under Element G: Teachers communicate effectively. It fell in the Developing category and was, "in the best way possible, communicate information to the students." This descriptor falls at a Developing level of performance because it describes effective communication without using a variety of methods.

Finally, six participants provided perceptions of Standard Four that fell under Element H: Teachers use a variety of methods to assess what each student has learned. Two of these fell in the Developing category. They were: "I'm aware of what they are receiving and learning" and "uses a variety of assessments." These descriptions fall at a Developing level of performance because they refer to the awareness or basic use of assessments. Three of the perceptions provided by participants fell in the Proficient category. These were: "I'm doing things that build 21st century skills," "using data to make sure we're teaching appropriately," and "using your data to group [students]." These fall within a Proficient level of performance because they refer to the use of assessment data and/or the use of 21st century knowledge in instruction. One of the descriptors fell in the Accomplished category. It was, "You're not just looking at data. You're looking at assessments and basing your practices off of assessments and data that you receive." This falls at an Accomplished level of performance because it describes the use of data to improve instructional practices.

Overall, none of the participants provided perceptions or descriptions of Standard Four that falls in the Distinguished level of performance. Five participants (22%) provided one descriptor of Standard Four that falls in the Accomplished level of performance. Nine participants (39%) provided one or two descriptors of Standard Four that fall in the Proficient level of performance. Four participants (17%) provided one or two descriptors that fall in the Developing level of performance. See Appendix H for a summary of participant perceptions of Standard Four as they relate to levels of performance.

Only one participant (4%) was able to provide descriptors of Standard Four that included four of the eight elements. One participant (4%) was able to provide descriptors that included three of the eight elements. Two participants (8%) were able to provide descriptors that included two of the eight elements. Eleven participants (48%) were able to provide descriptors that included one of the eight elements. Thirty-five percent of the participants were unable to provide any descriptors of Standard Four specific enough to identify any of the performance elements.

Standard Five: Teachers Reflect on Their Practice

When asked to identify Standard Five of the NCEES, none of the participants had any idea that Standard Five had to do with reflecting on one's practice. Once participants were told what Standard Five is, 48% described practices that fell under the analysis of student learning. Eight percent described practices that fell under professional growth. None of the participants described practices that fell under the third element of performance, functioning effectively in a complex and dynamic environment. Overall,

none of the participants seemed to have a full understanding of Standard Five. Table 9 summarizes these findings.

Table 9

Knowledge of Standards-Based Evaluation and Perceptions about Standards, Standard Five

Standard Five. Teachers Reflect on Their Practice							
	Analyze Student Learning	**Professional Growth**	**Function Effectively in a Dynamic Environment**	**Very Important Standard**	**Teachers Naturally Reflect**	**Not Enough Time to Reflect / Teachers Do Not Use Reflections**	**Not Observable**
P1				X			
P2					X		
P3	X						
P4						X	
P5		X					
P6		X					
P7	X						
P8							X
P9							X
P10	X						
P11					X	X	
P12							
P13					X		
P14	X						
P15	X					X	
P16	X				X		
P17	X						
P18	X						
P19	X						
P20						X	
P21	X						
P22	X						
P23					X		
Totals	**11**	**2**	**0**	**1**	**5**	**4**	**2**

Of the participants identifying a descriptor of Standard Five that fell under Standard Five, Element A: Teachers analyze student learning, none of the descriptors

were specific enough to be rated as a performance element. Rather, participants simply referred to reflecting on teaching practices. Of the two descriptors provided by participants of Standard Five that fell under Element B: Teachers link professional growth to their professional goals, both of them fell in the Proficient category. They fell at a Proficient level of performance because both were stated as simply participating in professional development. None of the participants provided a description of Standard Five, Element C: Teachers function effectively in a complex, dynamic environment. This element involves the use of research-based approaches to improve teaching and learning, and evaluating the impact of these approaches on student learning. See Appendix I for a summary of participant perceptions of Standard Five as they relate to levels of performance.

Overall, by looking at Appendices E through I, it is clear that participants as a group were quite limited in their ability to describe the instrument or provide examples of what each standard comprises. In addition, the majority of descriptions provided by participants were practices that fall within the Proficient or lower level of performance (Developing). It is not surprising that a lack of understanding and awareness of higher-level practices is prevalent considering the general lack of feedback participants receive from their experiences with the instrument (discussed later in this chapter).

Participants' Perceptions on the Fairness of Standards-Based Evaluation

When asked about the fairness of Standards One through Five, the standards-based portion of the NCEES, only four participants (17%) reported this portion of the evaluation tool as fair. Participant 14 said, "They do a good job of covering what you

need to be doing." Participant 16 reported, "The wording is consistent, the language is pretty consistent. You have a good understanding of what the standard is, what they're looking for in each area." Likewise, Participant 17 said, "I think it's kind of a snapshot of a well-rounded teacher so I don't see any flaws in it." Participant 21 simply said, "I don't think the standards themselves are unfair."

The remaining participants reported unfair qualities of the standards-based portion of the NCEES. Reasons stated by participants fall under the following categories:

1. Subjectivity
2. items that are not observable
3. observations only provide a snapshot
4. some teachers are at a disadvantage
5. VAD influence observation ratings
6. Poor communication or negative relationship with evaluator influences ratings

Appendix J provides a summary of the participants' views on the fairness of Standards One through Five of the NCEES.

In terms of subjectivity, Participant 3 argued, "I could see how it would be easy for it to be applied differently by different evaluators since it's so subjective." Participant 4 stated, "There's no way for the state to make sure that you're evaluating somebody the same way a principal at another school would evaluate somebody and because of that . . . there is no way to truly be fair." Participant 20 said, "[It's] very subjective, like different people can feel different ways about it." Participant 22 stated, "It's still subjective and not

useful for practice for me or with the students." Overall, 30% of participants felt that Standards One through Five are unfair due to subjectivity.

Twenty-six percent of participants felt that the standards-based portion of the NCEES is unfair due to items on the instrument that are not observable. For example, Participant 9 noted, "There are a lot of things that aren't observable in a 45-minute observation, such as . . . reflection." Participant 18 also argued that Standard Five is not observable. She also argued, "Parent relationships—unless you have a running phone log that you turn in . . . I don't know how you would know. Three observations is not going to give you that." Participant 23 said, "A lot of it is hard to see, and I hate getting a lower score just because you didn't see it."

Some participants felt that teachers who are not able or interested in accepting leadership roles are at a disadvantage with the NCEES, making it unfair. For example, Participant 1 argued, "I can't control what I'm going to get marked on [Standard One] because I'm a beginning teacher." Participant 6, referring to the Distinguished level of performance on some of the standards, said, "As a new teacher, nobody's going to listen to me so I never . . . know how to get to the next [level]." Participant 8 conveyed, "I don't think [they're] really appropriate for every single . . . teacher." Participant 9 said, "A lot of teachers aren't going to be leaders, but that doesn't mean they're not effective teachers. It's flawed, very flawed." Participant 10 commented,

> For a teacher to be evaluated as a leader, their strengths may be in the classroom . . . their strengths may not have anything to do with leadership in their building. I want to be in a classroom . . . and to me that's the dedication of a teacher and it shouldn't be scored down or looked down upon because a teacher might be a better teacher for not having her hands in everything.

Participants also stated that, depending on the subject they teach or the population of students they serve will also place them at a disadvantage on the NCEES. Participant 3 stated that some of the items "don't necessarily apply" depending on the subject you teach. Participant 11 said, "You're expected to grow every child in a year and not all children are alike, depending on what's going on in their lives that year." She went on to identify differences in parental involvement and support as factors that will contribute to a teacher's effectiveness ratings on the NCEES. Limited parental involvement due to language barriers, high mobility rates, and other factors are identified as unique factors that urban teachers face, as discussed in Chapter II.

Some participants felt that the standards-based portion of the NCEES is unfair because observations, in general, are just snapshots of a bigger picture. Participant 2 summarized her thought on this as, "[Evaluators] are seeing one chunk of what you do. They're seeing one portion of it. That's a disadvantage . . . it's only one piece of the entire puzzle." Participant 12 said, "[The observations] are snapshots. They're not seeing the whole picture, especially if you're only in the room for those three observations and then have to judge on those three observations. There are 180 days in the year and three days don't give you a big picture." Participant 15 said, "They are just measuring one moment in time…Ultimately, what is the end result of that measurement? Effective teaching is multifaceted and it's not one moment in time." Participant 22 argued, "During the time they're in here, I don't think they're going to be able to see the whole picture."

Poor communication, or a negative relationship, between the teacher and evaluator was another reason cited for the unfairness of the standards-based portion of the

NCEES. Specifically, two participants, or 8%, cited this reason. Participant 12 stated, "There's not been enough examples for me to know what I need to do . . . If there's any kind of personal conflict it shows up . . . what am I doing different . . . there's no communication on that. I haven't had clear feedback and I haven't had the support to fix it." Participant 19 said, "You're asked to give your input [but] I've learned to keep my mouth shut and just sign the paper because if you open your mouth you're going to get criticized."

Another stated reason that participants gave for the standards-based portion of the NCEES as being unfair was feeling that a teacher's VAD sometimes influences an evaluator's ratings on Standards One through Five. Two of the participants, or 8%, provided this reason. Participant 15 said, "I feel strongly that the test scores influence the observations."

Effects of Feedback on Professional Practices

One of the stated purposes of the NCEES is to help guide professional growth or development. Therefore, participants were asked about how they were last rated under each standard of the NCEES and what type of feedback they received from their evaluator. They were also asked to identify examples of how feedback under each standard has influenced or informed their professional practices. The purpose of these questions was to determine if the NCEES is actually influencing positive changes in professional practice.

Standard One: Teachers Demonstrate Leadership

Table 10 shows participants' ratings from their evaluators under Standard One. Thirteen percent of participants were rated as Developing. The majority of participants, 78% of them, were rated as Proficient in one or more elements of Standard One. Thirty-five percent of the participants were rated as Accomplished in one or more elements of Standard One. Only two participants (8%) received ratings of Distinguished under Standard One. Table 10 summarizes these findings.

When asked about feedback they have received under Standard One, 39% of participants could not recall receiving feedback under this standard. For example, Participant 19 stated, "Really, nothing much." Participant 20 stated, "I don't think any. Nothing that comes to mind." Most feedback that participants received was under leadership performance within the school. Thirty-five percent reported receiving positive feedback under this element. For example, Participant 11 said that her feedback was, "I do a good job, demonstrating [the standard] through leadership roles I hold." Participant 3 said, "It's been noted that I participate in staff meetings and vertical team meetings." Participant 4 said her feedback was, "I participate in PLC, lead staff development. I'm involved in helping new teachers to become acclimated to our school." Only Participant 14 (4%) received positive feedback under the element related to leadership within the classroom. He reported that his evaluator noted "that I take control of my class and that I lead it." Two participants (8%) received positive feedback under the element related to leadership within the profession. For example, Participant 9 noted, "I have only received things that are like, 'You did a nice job with our staff development.'"

Table 10

Teachers' Ratings under Standard One: Teachers Demonstrate Leadership

	Developing	Proficient	Accomplished	Distinguished
P1		X		
P2		X		
P3	X	X		
P4	X			
P5		X		
P6			X	
P7		X	X	
P8	X	X		
P9		X	X	
P10			X	
P11		X		
P12		X		
P13		X		
P14		X		
P15		X	X	
P16		X		
P17			X	X
P18			X	X
P19		X		
P20		X		
P21		X		
P22		X		
P23		X	X	
Totals	**3**	**18**	**8**	**2**

Only 13% of participants received any kind of constructive feedback under Standard One. One participant (4%) received constructive feedback under the element related to leadership in the classroom and two participants (8%) received constructive feedback under the element related to leadership in the school. For example, Participant 8 responded that her evaluator told her "to kind of step up a little bit more with joining

committees." Participant 16 said, "I am not considered accomplished because I could do more within the school." Participant 17 said, "He talked to me about facilitating leadership within my classroom, which is building leadership with my students, leading within their teams." None of the participants received constructive feedback under the element related to leadership in the profession. Additionally, none of the participants received any feedback, positive or constructive, under the element related to advocacy for schools or students or the element related to ethical behaviors.

Only two participants could provide an example of how feedback under Standard One has helped to influence their professional practices. Participant 16 said, "Using that feedback, I did join a lot more committees this year and I have moved on to doing more community-based assignments. I'm taking this year's feedback to move more towards district things as well." Participant 17 stated, "I took some of his suggestions and tried to integrate that into my teaching." Participants 5 and 13 reported that professional resources and graduate school were sources that help them to alter their practices in the classroom.

Table 11 presents findings related to feedback under Standard One. The general lack of constructive feedback, despite the fact that every participant had room to grow in at least one element of Standard One, supports O'Donovan's (2011) claim that teacher evaluation is not necessarily going to lead to improvements in teacher performance. This is particularly true if teachers do not receive relevant, constructive feedback to help them improve. This lack of feedback is prevalent in all standards of the NCEES for the participants in this study.

Table 11

Feedback under Standard One: Teachers Demonstrate Leadership

	Positive Feedback					Constructive Feedback						
	Leadership in Classroom	**Leadership in School**	**Leadership in Profession**	**Advocacy**	**Ethics**	**Leadership in Classroom**	**Leadership in School**	**Leadership in Profession**	**Advocacy**	**Ethics**	**Examples of Effect on Practices**	**No Feedback Provided**
P1		X										
P2		X										
P3		X										
P4		X										
P5		X										
P6			X									
P7												X
P8							X					
P9			X									
P10												X
P11		X										
P12		X										
P13												X
P14	X											
P15												X
P16							X				X	
P17						X					X	
P18												X
P19												X
P20												X
P21		X										
P22												X
P23												X
Total	**1**	**8**	**2**	**0**	**0**	**1**	**2**	**0**	**0**	**0**	**2**	**9**

Standard Two: Teachers Establish a Respectful Environment for a Diverse Population of Students

When asked how they were last rated under Standard Two, one participant had been rated as Developing and one participant had been rated as Distinguished. Another participant could not recall what ratings they received under Standard Two. Most of the participants were rated as Proficient and/or Accomplished under Standard Two. Seventy percent of the participants stated that they had received ratings of Proficient in one or more elements of Standard Two. Fifty seven percent had received ratings of Accomplished in one or more elements of Standard Two. Table 12 summarizes participant ratings under Standard Two.

As Table 13 illustrates, sixty-five percent of the participants could not recall receiving any feedback from their evaluators under Standard Two. For example, Participant 7 stated, "I honestly don't remember any type of feedback." Participant 8 noted, "I don't think I've had any major feedback for that one." Participant 12 said, "I don't remember ever having something actually written."

Six of the participants (26%) received positive feedback under the element related to providing a positive and nurturing relationship and environment for students. Participant 14 said that his evaluator "liked the way I treated kids . . . my classroom was always a very welcoming environment." Participant 18 said her evaluator said, "I build strong relationships with every student." Participant 1 said her evaluator told her "that I show good rapport with my students, that I make them feel comfortable." Participant 2 stated that she was told, "that I promote a positive environment in my classroom."

Table 12

Teachers' Ratings under Standard Two: Teachers Establish a Respectful Environment for a Diverse Population of Students

	Developing	Proficient	Accomplished	Distinguished	Cannot Remember
P1			X		
P2		X			
P3	X	X			
P4		X	X		
P5		X			
P6			X		
P7		X	X		
P8		X	X		
P9		X			
P10		X			
P11		X	X		
P12		X	X		
P13			X		
P14		X	X		
P15			X		
P16		X			
P17					X
P18			X	X	
P19		X	X		
P20		X			
P21		X			
P22			X		
P23		X			
Totals	**1**	**16**	**13**	**1**	**1**

Table 13

Feedback Under Standard Two: Teachers Establish a Respectful Environment for a Diverse Population of Students

	Positive Feedback					Constructive Feedback						
	Positive, Nurturing Relationship	**Embrace Diversity**	**Treat Students as Individuals**	**Adapt Teaching for Students**	**Work Collaboratively with Families**	**Positive, Nurturing Relationship**	**Embrace Diversity**	**Treat Students as Individuals**	**Adapt Teaching for Students**	**Work Collaboratively with Families**	**Examples of Effect on Practices**	**No Feedback Provided**
P1	X											
P2	X											
P3												X
P4	X											
P5												X
P6	X											
P7												X
P8												X
P9												X
P10												X
P11												X
P12							X					
P13												X
P14	X											
P15												X
P16			X								X	
P17											X	X
P18	X											
P19												X
P20												X
P21												X
P22												X
P23												X
Total	**6**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**15**

One participant, Participant 16, received positive feedback under the element related to treating students as individuals. She said her evaluator noted, "that I do a good job of understanding my students' needs. I recognize them as individuals." Only Participant 12 received any type of constructive feedback under Standard Two and that was under the element related to embracing diversity in the school community and in the world. She reported, "the creating lessons for a diverse population . . . I don't remember specific feedback other than I needed to work on it." None of the participants received any type of feedback related to the elements of adapting teaching for the benefit of students with special needs or working collaboratively with families. Not a single participant could provide an example of feedback that has influenced their practices under Standard Two. Participant 5 noted, "I can't think of anything, any feedback that's been given to me that would cause me to change practice, which is awful but true." Participant 19 stated, "I think where I get my most feedback is what I see that works and doesn't work." Again, it is noteworthy that despite the exponential growth of demand for accountability of teachers, participants in this study had a virtual absence of support in terms of feedback to actually improve.

Standard Three: Teachers Know the Content They Teach

Under Standard Three, most participants had been rated as Proficient and/or Accomplished by their evaluators. None of the participants had been rated as Developing under Standard Three. Seventy-eight percent of participants had been rated as Proficient in one or more elements of Standard Three. Forty-four percent of participants had been rated as Accomplished in one or more elements of Standard Three. Only two participants

(8%) had been rated as Distinguished in one or more elements of Standard Three. Table 14 depicts participant ratings under Standard Three.

Table 14

Teachers' Ratings under Standard Three: Teachers Know the Content They Teach

	Developing	Proficient	Accomplished	Distinguished
P1			X	
P2		X		
P3		X		
P4		X		
P5			X	
P6		X	X	
P7		X	X	
P8			X	
P9		X		
P10			X	X
P11		X		
P12		X		
P13		X		
P14		X		
P15		X	X	
P16		X		
P17			X	X
P18		X	X	
P19		X	X	
P20		X		
P21		X		
P22		X		
P23		X		
Totals	**0**	**18**	**10**	**2**

Over half of the participants (52%) could not recall receiving any feedback under Standard Three. For example, Participant 8 stated, "I haven't really received any feedback." Participant 22 stated, "Nothing that I can remember." Three participants

(12%) received positive feedback under the element related to instruction being aligned with the curriculum and two participants (8%) received constructive feedback under this element. The positive feedback was, "that I follow the common core guidelines and the expectations for the county," (Participant 2) "The only real feedback I get on Standard Three is that my principal truly believes that I absolutely hands down know the curriculum," (Participant 5) and "that I do a great job of incorporating common core standards in my classroom" (Participant 16). The corrective feedback was given to Participants 9 and 18: "that I don't know the common core" (Participant 9) and "feedback saying that to continue the study of common core . . . I am teaching the curriculum . . . but that I can continue to grow in terms of the depth of knowledge" (Participant 18).

Four participants (17%) received positive feedback under the element related to knowing the content appropriate to their teaching specialty. Participant 1 stated that her evaluator noted, "that I know the content in which I'm teaching." Participant 7's evaluator noted, "that I know my content and can teach it." Participant 14 reported that his evaluator told him, "I know the content area." One participant (4%) received constructive feedback under this element. That was Participant 12, whose evaluator told her that she needed to be "more comfortable with the content before teaching it."

Eight percent received positive feedback that was too generic to fall under one of the specific elements under Standard Three. Participants 3 and 10 could only recall being given positive feedback, but could not recall any specifics. None of the participants received any type of feedback, positive or constructive, under elements related to understanding the interconnectedness of content areas or making instruction relevant to

students. None of the participants could provide any examples of feedback that they have been able to utilize to inform or influence their professional practices. For example, Participant 15 said, "To be honest, we sit down, we look at it . . . I've signed off on it . . . Do I change practices because of it? No." This reference to an automated and nearly thoughtless process is consistent with what Winters (2012) describes as a "rubber stamp approach" to teacher evaluation. Table 15 summarizes these findings.

Standard Four: Teachers Facilitate Learning for Their Students

Under Standard Four, most participants had been rated as Proficient or Accomplished in one or more elements. Only two participants (8%) had received ratings of Developing under Standard Four. Sixty-one percent of participants reported having ratings of Proficient in one or more elements of Standard Four. Likewise, 61% reported having ratings of Accomplished in one or more elements of Standard Four. Only 2 participants (8%) had former ratings of Distinguished under Standard Four. Table 16 summarizes the participants' ratings under Standard Four.

The majority of participants (53%) could not recall receiving any feedback under Standard Four. For example, Participant 3 said, "I still haven't gotten much back on that from any area." Participant 19 responded "generally nothing specific" about feedback related to practices under Standard Four. Participant 21 said, "I can't recall any [feedback] specifically."

Table 15

Feedback under Standard Three: Teachers Know the Content They Teach

	Positive Feedback					Constructive Feedback						
	Instruction is Aligned	**Teachers Know Content Appropriate to Specialty**	**Recognize Interconnectedness**	**Make Instruction Relevant**	**Generic, Does Not Fit Under Elements**	**Instruction is Aligned**	**Teachers Know Content Appropriate to Specialty**	**Recognize Interconnectedness**	**Make Instruction Relevant**	**Generic, Does Not Fit Under Elements**	**Examples of Effect on Practices**	**No Feedback Provided**
P1		X										
P2	X											
P3					X							
P4												X
P5	X											
P6												X
P7		X										
P8												X
P9						X						
P10					X							
P11												X
P12							X					
P13												X
P14		X										
P15												X
P16	X	X										
P17												X
P18						X						
P19												X
P20												X
P21												X
P22												X
P23												X
Total	**3**	**4**	**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**0**	**12**

Only four participants (17%) reported receiving positive feedback under Standard Four. Those four participants reported receiving feedback in only one of the following elements: knowing appropriate levels of development, using a variety of instructional methods, integrating technology, and using a variety of methods to assess students. Participant 10, for example, stated that her evaluator made comments about "data and evaluation and assessments and how I use that in class . . . encouragement to continue." Participant 2 said that her evaluator noted, "It's evident that I meet the needs of the students, how I had differentiated." Participant 18's evaluator noted how she "consistently used technology in an effective way."

Table 16

Teachers' Ratings under Standard Four: Teachers Facilitate Learning for Their Students

	Developing	**Proficient**	**Accomplished**	**Distinguished**
P1			X	
P2			X	
P3	X	X		
P4		X		
P5		X		
P6			X	
P7		X	X	
P8			X	X
P9		X	X	
P10			X	
P11			X	
P12	X	X		
P13			X	
P14		X		
P15		X	X	
P16		X	X	
P17			X	X

Table 16

(Cont.)

	Developing	Proficient	Accomplished	Distinguished
P18			X	
P19		X	X	
P20		X		
P21		X		
P22		X		
P23		X		
Totals	**2**	**14**	**14**	**2**

Three participants (12%) reported receiving constructive feedback under the element related to understanding the developmental levels of students. Participant 12, for example, said that her evaluator noted, "that I am not differentiating." Participant 20's evaluator told her to "try to differentiate more." Twelve percent reported receiving constructive feedback under the element related to having students learn in teams. For example, Participant 5 noted that her evaluator told her, "The only way I'm grouping kids is in tribes" and asked, "Are you not grouping them by data?" Participant 22's evaluator told her to try "dividing the students into groups and encouraging hands-on with the students." One participant (4%) received constructive feedback related to planning instruction appropriate for students (Participant 12: "I'm not using data in the classroom") and one participant received constructive feedback related to using a variety of methods to assess what students have learned (Participant 8: her evaluator is "wanting to see more data-driven instruction, especially in science").

Only three participants (12%) could provide an example of receiving feedback that has influenced their instructional practices. For example, Participant 2 said, as a

result of feedback, “I use a lot more rubrics so the students can go through and evaluate themselves and know the expectations.” Participant 6 said that she is using different levels of text as a result of feedback. She reported, “like the articles . . . on 3D printing . . . [students] are each reading a different level of text about the main topic.” Participant 20 noted, “We talked about text complexity and choosing exemplar texts that DPI has put out.” Other participants, such as Participants 3, 7, 8, 9 and others, could not recall any examples of feedback that has influenced their practices. Participant 12 voiced frustration over the lack of feedback. She said, “I was not given the definition of what I was being asked [to do].” Participant 15 was asked to differentiate for students, but she said, “That’s not really going to influence me at this point . . . I cannot say that it will change what I do. I think it’s just that one moment in time, so I’m not going to take it so seriously.” Overall findings of participants’ feedback under Standard Four are summarized in Table 17.

Standard Five: Teachers Reflect on Their Practice

Under Standard Five, none of the participants were rated as Developing. Fifty-seven percent of the participants reported being rated as Proficient under one or more elements of Standard Five. Sixty-one percent of participants reported being rated as Accomplished under one or more elements of Standard Five. Four participants (17%) reported being rated as Distinguished under one or more elements under Standard Five. One participant could not recall how she had been rated under Standard Five. Table 18 shows how participants were rated under Standard Five.

Table 17

Feedback under Standard Four: Teachers Facilitate Instruction for Their Students

	Positive Feedback								Constructive Feedback									
	Know Appropriate Levels of Development	**Plan Appropriate Instruction**	**Use a Variety of Instructional Methods**	**Integrate Technology**	**Develop Critical Thinking & Problem Solving**	**Help Students Work in Teams**	**Communicate Effectively**	**Use a Variety of Methods to Assess**	**Know Appropriate Levels of Development**	**Plan Appropriate Instruction**	**Use a Variety of Instructional Methods**	**Integrate Technology**	**Develop Critical Thinking & Problem Solving**	**Help Students Work in Teams**	**Communicate Effectively**	**Use a Variety of Methods to Assess**	**Examples of Effect on Practices**	**No Feedback Provided**
P1																		X
P2	X																X	
P3																		X
P4			X															
P5														X				
P6														X			X	
P7																		X
P8																X		
P9																		X
P10								X										
P11																		X
P12									X	X								
P13																		X
P14																		X
P15									X									
P16																		X
P17																		X
P18				X														
P19																		X
P20									X								X	
P21																		X
P22														X				
P23																		X
Totals	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**3**	**1**	**0**	**0**	**0**	**3**	**0**	**1**	**3**	**12**

Table 18

Teachers' Ratings under Standard Five: Teachers Reflect on Their Practice

	Developing	Proficient	Accomplished	Distinguished
P1		X	X	
P2			X	
P3		X		
P4		X		
P5			X	X
P6		X	X	
P7		X	X	
P8			X	X
P9		X		
P10			X	
P11			X	
P12		X	X	
P13				
P14		X		
P15			X	X
P16			X	
P17			X	
P18			X	X
P19		X	X	
P20		X		
P21		X		
P22		X		
P23		X		
Totals	**0**	**13**	**14**	**4**

Ninety-one percent of participants could not recall receiving any type of feedback under Standard Five. For example, Participant 3 said, "I don't recall any" and Participant 7 said, "None, that I can remember." Participant 9 argued, relative to Standard Five, "The nature of these evaluations is I'm going to watch you for 45 minutes and then I'm going to evaluate you and guess." This comment highlights the perception of some of the participants that not all elements of the NCEES are observable. It also reflects the

experiences that some participants have with the evaluation process that is automatic and void of thoughtful feedback. This supports Darling-Hammond et al.'s (2012) claim that teacher evaluation does little to promote the professional development of teachers.

One participant (4%) reported receiving positive feedback under the element related to linking professional growth to professional goals. That was Participant 2 and she reported that her evaluator noted, "that I participate in professional development and vertical teams and those kinds of things. Just general." Participant 8 (4%) reported receiving constructive feedback under the element related to analyzing student learning. She noted, "They would like to see more [reflection] . . . to keep better documentation of why I'm doing what I'm doing, to reflect about it." None of the participants could recall any feedback under Standard Five that has influenced their professional practices, though Participant 8 did say she was "trying to record for documentation purposes," but then noted, "I wouldn't say it's improved my practices of reflecting." Table 19 summarizes participant responses related to feedback under Standard Five.

In summary, reasons why standards-based feedback does not influence professional practices may be classified into five general themes. These are:

1. feedback that is too generic to use
2. feedback that is descriptive in nature (simply describes what is seen, but does not judge)
3. feedback that is positive in nature and validates current practices
4. feedback that is constructive in nature, but ignored
5. complete lack of feedback

Table 19

Feedback under Standard Five: Teachers Reflect on Their Practices

	Positive Feedback			Constructive Feedback			Examples of Effect on Practices	No Feedback Provided
	Analyze Student Learning	Link Professional Growth to Professional Goals	Function Effectively in Dynamic Environment	Analyze Student Learning	Link Professional Growth to Professional Goals	Function Effectively in Dynamic Environment		
P1								X
P2		X						
P3								X
P4								X
P5								X
P6								X
P7								X
P8				X				
P9								X
P10								X
P11								X
P12								X
P13								X
P14								X
P15								X
P16								X
P17								X
P18								X
P19								X
P20								X
P21								X
P22								X
P23								X
Totals	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**21**

The first three reasons are simple to understand why no changes in professional practice would occur; there is either a lack of substance to the feedback or it simply summarizes or validates the practice already being utilized. The diminutive amount of constructive feedback actually received by participants that is ignored may be due to the lack of relationship with the evaluator or the perception that the feedback is not accurate. Recall from Chapter III the human tendency to be threatened by constructive feedback if one of these two conditions exist (Feys et al., 2008; Leary, 2007; Sedikides & Green, 2004). The most common reason, however, reported by the participants in this study to indicate why the evaluation process does not lead to improvements in professional practice is due to the final reason, which is the total absence of constructive feedback during the teacher evaluation process.

Knowledge of VAD and Standard Six of the NCEES

Fourteen out of the 23 participants (61%) understood that VAD is based on student growth. The remaining 9 participants (39%) were not sure on what VAD is based. Some participants, like Participant 4, thought that growth might be part of VAD, but was unsure. She commented, "Every teacher's value is based on her kids' scores. That's all I remember . . . To truly be fair, you have to take how much they have grown in the year from where they started. Is that included? See, I don't know." Participant 20 thought that VAD included growth, but also included students' proficiency rates. She said, "I know it has something to do with test scores. I think it's partially how many of your kids are proficient and partially how many have made growth." Other participants did not have any idea how to explain VAD. For example, Participant 19 said, "I just know how the

students do is related to how I will be graded." Participant 16 admitted, "I don't remember the specifics of it." Participant 12 said, "I understand that I'm being scored on it, that there is a formula out there . . . but we are still confused as to how that happens and where it comes from."

Most of the participants who understood that VAD reflects student growth, communicated a lack of confidence in their knowledge of VAD. In fact, only one participant (Participant 5) seemed to have a thorough understanding of VAD. She explained VAD as "an incredibly complicated algorithm created by SAS, who has taken not just my year with the kids but all of the years and created a projected growth that, based on the past performance of this child, this kid should be able to get to a certain number and if they don't meet that number, there's a loss involved." However, many of the participants communicated that they did not feel they had a thorough knowledge of or confidence level with VAD. For example, Participant 1 said, "I don't really have a lot of understanding." Participant 7 acknowledged, "I don't know much about it. I know that it's some formula they put together and I think it's for Standard 6, right?" Participant 8 admitted, "I don't have a very good understanding of VAD." Table 20 summarizes the participants' understanding of VAD.

Table 20

Participants' Knowledge of VAD

	Based on Growth (Accurate Understanding)	Involves Projections on Performance (Partial Understanding)	Based on Proficiency (Inaccurate Understanding)	Not Sure (Lack of Understanding)	Describes Own Understanding as Limited
P1	X				X
P2	X				
P3	X				
P4				X	X
P5	X				
P6	X				X
P7				X	X
P8				X	X
P9		X			X
P10	X				
P11	X				X
P12		X			X
P13	X				
P14	X				
P15	X				X
P16				X	X
P17		X			
P18	X				
P19				X	X
P20			X		X
P21	X				X
P22	X				
P23	X				
Totals	**14**	**3**	**1**	**5**	**13**

When asked about their own VAD, only two participants, or 8%, quantified their VAD status with Level 1 through 5. Participant 5 stated that she is a Level 2 teacher. Participant 17 said that she is a Level 4 teacher with her seventh grade students and a

Level 5 teacher among her eighth grade students. Nine participants, or 39%, could describe their VAD. For example, Participant 22 said her VAD was "negative 1." Participant 1 stated, "Math was good but science was down a little bit." Participant 7 said, "I remember my data being on the high end of green. I know that I work well with the low kids compared to the high kids." Participant 8 noted, "It wasn't great, but it wasn't horrible. I remember it just being kind of mediocre." Participant 11 recalled, "I wasn't in the green and it was barely not [in the green]." Participant 16 commented, "I was right in the middle. I grew my advanced learners the most." Participant 18 remembered being "stronger with the middle group of students." She further recalled, "My lower students have been the students that I'm not able to meet growth with." Participant 21 said, "My growth was poor." Finally, Participant 23 said, "I showed growth in everything but math."

The remaining 53% or 12 participants could not provide any details about their VAD. In fact, a couple of participants conveyed that their VAD is unimportant. For example, Participant 3 admitted, "I never took the trouble to find out. I think most teachers just went, 'This is too much trouble' and I don't know that anybody's looked at it." Participant 22 acknowledged, "I don't know much about [VAD]. I don't put a lot of weight on that." Table 21 summarizes participants' knowledge of their own VAD.

Table 21

Participants' Knowledge of Their Own VAD

	Identified Assigned Level (1–5)	Identified Color	Knew Some Specifics of Own VAD	Identified Specifics of Student Groups	Did Not Know Details of Own VAD
P1			X		
P2					X
P3					X
P4					X
P5	X				
P6					X
P7		X			
P8			X		
P9					X
P10					X
P11		X			
P12					X
P13					X
P14					X
P15					X
P16				X	
P17	X				
P18				X	
P19					X
P20					X
P21			X		
P22			X		
P23			X		
Totals	**2**	**2**	**5**	**2**	**12**

Participants' Perceptions on the Fairness of VAD

The majority of participants (92%) has perceptions that VAD is unfair and should not be included in the NCEES. The participants provided a variety of reasons that VAD is unfair. The most common reasons provided fall under the concept that too many factors

outside a teacher's control affect VAD. Forty-eight percent of the participants' conveyed this opinion. For example, Participant 1 stated, ". . . some kids just don't test well." Participant 8 stated that her problem with VAD is, "There are so many factors that affect my data." Participant 10 described, "There are some years they don't have the support at home, they've never been nurtured and taught from home and I think a lot of accountability for what should be happening with mom and dad is dumped on a teacher." Participant 18 argued, "It does not account for anything going on in that child's life, whether their parents are getting a divorce or they didn't have breakfast." Participant 4 asked, "Is there a place where a teacher can say, 'These are the reasons for lack of growth?'" Participant 3 pointed out that in her district, each school is responsible for calculating the percentage of time each student is assigned to the teacher and that, because the guidelines for this varies from school to school, it is an extraneous factor that makes VAD unfair. She described, ". . . having to put in percentage of time, that was a mess and I had to do it for two schools last year and the way the two schools were doing it were completely different . . . so, at the moment, I still think it's not very accurate data." This inconsistency in how VAD is formulated illustrates the argument explained by Barile (2013) and Papay (2012) that a derived numerical value is not necessarily objective and without flaw.

Seventeen percent of participants felt that VAD is unfair because they believe that teachers with low-performing students are at a disadvantage. Participant 20 stated, "You theoretically should see growth among all students from year to year but if they didn't know last year's content, more than likely, they're not going to be proficient in this year's

content." Participant 22, who conveyed support for VAD and is not included in the 17%, still expressed concern about students who may be lower performing. She said, "Hopefully, they're balancing out whether they're students [in special education] because I know that all classrooms don't have the same mix of students in them." Participant 23 argued, "If you're in a tough school with tough children and 60% of your children are not proficient when you get them at the beginning of the year . . . I mean, I only grew half of them and that doesn't look good. I think that's the part that's frustrating." Participant 11 commented that it is "not fair to a teacher that gets lower classes, even though they say it's fair" and Participant 13 said, "you have teachers that teach different things, teachers who are teaching just the advanced kids, so I think it's hard to do it in a way that is fair across the board."

Conversely, two participants (8%) argued that teachers are at a disadvantage with VAD if they teach high-performing students. Participant 5 said, "I'm working with 99, 98, 97 percentile kids who have nowhere to grow but I'm being held accountable for the fact that they don't grow." Participant 14 explained, "It's hard to maintain that 98 percentile or even show growth and then [teachers] are still penalized for that . . . Over so many years if you maxed out your growth, I don't think you should be penalized."

One participant (4%) communicated that VAD is unfair but did not provide a reason as to why it is unfair. One participant (4%) stated that VAD doesn't seem accurate because it fluctuates so much from year to year and one participant (4%) simply described VAD as "too subjective." The two participants (8%) who communicated support for VAD did so with some hesitation. Participant 6 stated, "Well, I'm actually not

sure how it works. I have been told my numbers weren't bad. Maybe I would feel differently if they weren't good." Participant 22, who conveyed support for VAD despite her reservations about students with special needs adversely affecting teachers' VAD, said, "As long as it's done by scientists in that area and it's accurate, then I think it's good to include . . . I'm trusting that those people who have come up with the formulas have done it accurately and fairly." Table 22 summarizes participant perceptions on the fairness of VAD.

Table 22

Participants' Perceptions of Fairness of VAD

	FAIR	**VAD IS NOT FAIR**						
	Fair, But Has Reservations	**Too Many Factor Affect VAD**	**Low Students Affect VAD**	**High Students Affect VAD**	**Too Much Emphasis on One Measure**	**Too Inconsistent**	**Subjective**	**No Specific Reason Given**
P1		X						
P2		X						
P3		X						
P4		X						
P5					X			
P6	X							
P7							X	
P8		X						
P9		X						
P10		X						
P11			X					
P12					X			
P13			X					
P14				X				
P15								X
P16						X		
P17		X						

Table 22

(Cont.)

	FAIR	VAD IS NOT FAIR						
	Fair, But Has Reservations	Too Many Factor Affect VAD	Low Students Affect VAD	High Students Affect VAD	Too Much Emphasis on One Measure	Too Inconsistent	Subjective	No Specific Reason Given
P18		X						
P19		X						
P20			X					
P21		X						
P22	X							
P23			X					
Totals	**2**	**11**	**4**	**1**	**2**	**1**	**1**	**1**

Participants were also asked how they were rated under Standard Six on the NCEES. None of the participants exceeded expectations on Standard 6. Two participants (8%) met expectations on Standard Six. Two participants (8%) did not meet expectations on Standard 6. The remaining 84% of participants (19 of the participants) could not recall how they were rated on Standard 6. Table 23 summarizes these findings.

Table 23

Participant Ratings on Standard Six

	Exceeded Expectations	Met Expectations	Did Not Meet Expectations	Cannot Recall Rating
P1		X		
P2				X
P3				X
P4				X

Table 23

(Cont.)

	Exceeded Expectations	Met Expectations	Did Not Meet Expectations	Cannot Recall Rating
P5			X	
P6				X
P7				X
P8				X
P9				X
P10				X
P11			X	
P12				X
P13				X
P14				X
P15		X		
P16				X
P17				X
P18				X
P19				X
P20				X
P21				X
P22				X
P23				X
Totals	**0**	**2**	**2**	**19**

Trends in Participants' VAD and Their Reflections on the Usefulness of VAD

Participants were asked how their VAD compared across years. Only six participants (26%) could recall any information about trends in their VAD across years. One participant reported that her VAD has been improving over time. Three participants reported recent declines in their VAD. Two participants reported consistent VAD over time. The remaining 74% of participants could not provide any information about their VAD over time.

Only two participants, or 8%, were able to identify any sort of reflection from her VAD. Participant 6 said,

> It made sense to me because I had a super strong bond with the high eighth graders because they had my humor and we just got along really well. Then, the next year he said that I was better with the lower students and that made sense because I remember making a really strong bond with my homeroom . . . I knew I was connecting differently but it was interesting to see it impact scores. If they're connected more to you, they're going to try harder.

Participant 18 said, "I knew I needed to focus my efforts more on small group instruction with my lower students."

Fifty-two percent of participants could not make any reflections on their VAD because they could not recall enough details about the data. One participant (4%) noted that the school, as a whole, reflected on the science data and changed the master schedule as a result to allow for more instructional time in science. The remaining 35% of participants indicated that they are unable to make beneficial reflections on their VAD because it provides limited information or because they do not find it useful for that purpose. Participant 7 clarified, "Every year, I try to fine tune all my strategies, just

because of my own thinking." Participant 8 said, "I didn't need that data to remind myself that I needed to improve." Participant 10 admitted, "I don't think it changes . . . the way I teach." Participant 11 argued, "You get frustrated because you've worked and worked and it doesn't make any difference and it's still based on a standardized test score and if you look at it, test scores have stayed about the same for 20 years no matter what we do." Participant 15 said, "It only takes a few [students] to move that [VAD] and that's not fair." Finally, Participants 17 and 21 agreed that VAD has limited value for the purpose of reflecting. Participant 17 summarized it by saying, "VAD doesn't give you specific feedback as far as 'here is what you could do to move those kids.'" Participant 21 said, "It's hard to pinpoint one particular thing or multiple things that would help increase student growth or learning."

Appendix K provides a summary of participant responses related to the usefulness of VAD for reflective purposes. Overall, the findings support one of the arguments provided by Green et al. (2012) against the use of VAD described in Chapter II. Multiple years of data are needed. However, most teachers do not have multiple years of data and only two participants (8%) from this study had VAD that was consistent over time.

Participants' Recommendations for Improvements in Teacher Evaluation

Participants offered numerous suggestions for improving the teacher evaluation process in North Carolina. The suggestions fall into the following categories:

1. Devote more time to a quality process
2. Omit VAD from the instrument.

3. Provide better feedback to teachers so they have a better understanding of how to improve.
4. Make the instrument shorter
5. Omit peer observations from the process
6. Use multiple evaluators, including outside evaluators.
7. Eliminate the levels of performance (performance ratings) and provide written feedback only.
8. Provide more training to teachers and/or evaluators on the instrument.
9. Use multiple formats.

One of the most commonly stated suggestions for improvement is to devote more time to a quality process. Participants describe a quality process as one that allows the evaluator to spend sufficient time in the classroom and with the teacher during the discussion of the evaluation. Nine participants, or 39% provided suggestions that fell in this category. For example, Participant 11 said, "I think they can certainly evaluate on a more informal basis, if we trust our principals and assistant principals…They don't get to spend enough time doing things that could be more effective." Participant 13 suggested, "Maybe they could actually spend quality time with teachers because it just seems like such a rushed process." Participant 17 described this issue by saying, "We have one principal and he has to do 36 evaluations, so when it comes time for my conference he's like, 'Well, you know the deal, so we're going to look over it really quick and if you have any questions, let me know,' so maybe you don't get the attention you need with it." Participant 18 explained, "There's not enough emphasis put on it. It's

seen as, ‘we have to do that’ rather than ‘this is best for your teaching.’” Participant 20 said, “[Evaluators] were just trying to get it done. ‘Gotta meet the deadline’ as opposed to it being meaningful . . . Come around more often.” Participant 22 noted, “[Evaluators] don’t really have time to meet with teachers. I was evaluated a few weeks ago and I still haven’t been called to sign off on anything or to discuss it.”

Another thirty-nine percent of participants suggested that VAD be omitted from the NCEES or improve VAD methods. Participant 2 noted, “There are too many outside factors playing into [VAD].” Participant 3 said, “It seems as fair as you can make it, short of Standard 6, which they need to get a little better organized before we use that data.” Participant 5 commented, “The VAD . . . it does a really lovely job of rank ordering teachers . . . but I don’t think it lets anybody know how effective I am.” Participant 7 argued, “It’s hard to measure a teacher against a teacher when they don’t have the same kids.” Participant 11 suggested, “If we took the money that we spend on all the people that analyze that data and put it back in the classrooms, we’d get a lot more bang for our buck…less data, less people at central offices that never see students in the classrooms.” Participant 14 stated, “This goes back to the student growth. If there is some other way besides just a standardized test . . . but what else would you use? I don’t know.” Participant 15 argued, “It’s just the way it’s done. Test scores shouldn’t influence the [observations]. That should stand alone. Test scores need to have no influence whatsoever on the other standards.” Participant 21 said, “With Standard 6, I think it should come from a range of assessments, not just a single end-of-the-year assessment.” Participant 23 argued that working in a Title 1 school puts teachers at a disadvantage in

terms of VAD. She said, “I don’t like the VAD part. It’s just so difficult when you have tough schools, when you have Title 1 schools.”

Thirty percent of participants suggested that the teacher evaluation process could be improved by improving the feedback that teachers receive during the process. For example, Participant 5 said, “I would like to see clearer guidelines on how you move from one [performance level] to the next, like what does that really look like in a classroom?” Participant 8 said, “Just more feedback . . . evaluators come in and they might just write ‘great job’ but they don’t really give you a lot of feedback to improve instruction.” Participant 10 argued, “The feedback . . . is just a general statement. That instrument keeps them from [being] available. It would be better if they could just be in the classroom and it be a less formal instrument.” Participant 13 said something similar. She said, “Maybe if they could actually spend quality time with teachers because it just seems like such a rushed process. Maybe if they could . . . push us to the next level, but it’s more of . . . writing down what you already do.” Likewise, Participant 17 shared, “It’s a revolving door of due dates . . . sometimes it just gets to crunch time and you don’t get as much feedback as you might want or need.” Participant 21 suggested, “They could give more specific feedback to teachers. The feedback is there but most of it is not very specific.”

Twenty-two percent of participants suggested that the evaluation instrument should be shorter for classroom observations. Participant 1 said, “There is a lot done . . . that may not be observed but it’s on the evaluation.” Participant 5 said, “I would like to see it streamlined.” Participant 9 suggested, “Make it shorter. This think is way too long,

it's way too big . . . It needs to be similar to feedback we give our students . . . if we give them a half an hour monologue about how they can improve, they're not going to get it and [it's] the same with teachers." Participant 10 argued, "It's so time consuming . . . That instrument keeps them from [being] available." Participant 18 said, "There are too many levels, too many questions, too many areas . . . That's just too much . . . If they were to make it a shorter process or not so tedious, it might be more helpful to teachers."

An additional 22% of participants felt that the evaluation process could be improved if the peer observation component was omitted. Participant 3 stated, "I don't know that the peer evaluation has much value. You don't want to upset your peers by marking them low so I can't imagine that anybody is ever going to of out of their way to be critical." Participant 8 said, "I've done peer observations and I don't know half the stuff on there. It's very uncomfortable . . . you didn't even want to focus on the negative because you don't want to have that type of relationship with your coworker." Participant 12 said, "Peer observations—we're more lax and nicer so I don't think that [is] effective." Participant 15 said, "We've been asked to observe our peers and we're not as proficient in it as maybe we should [be]. It's hard for me to . . . observe someone." Finally, Participant 20 commented, "Peer evaluators, I don't know that they mark bad things. They're just trying to get it done."

Twenty-two percent of participants also suggested that multiple evaluators would improve the teacher evaluation process. Participant 8 suggested, "Have other people make observations, [that] would be beneficial." Participant 12 noted "an outside person coming in to do the evaluations" would make it more objective. Participant 18 said,

"Even if it was an outside evaluator who came in just for evaluations, that would maybe be an unbiased way. Then maybe people wouldn't feel like it was so subjective, so they might be more open to the feedback." Participant 20 suggested, "Maybe having multiple people rate, that would take some of the subjectivity out of it." Participant 22 also suggested, "Different evaluators . . . could be more helpful."

Twelve percent of participants suggested that omitting the ratings on level of performance would make the evaluation process better and conveyed a desire to have verbal feedback without the ratings. Participant 6, for example, said, "I would love for it just to move to written feedback. It just seems like a paperwork requirement. I grapple with being a number or a checkbox." Participant 9 described the evaluation system in another state, "They would just write a paragraph" and argued, "There was no long evaluation system and the kids got to know [administrators] so there was more administrative support for teachers and students." Participant 22 suggested, "Just maybe list out the things that the teachers are doing well, and the suggested improvements, and not see ratings."

Another 12% of participants suggested that the evaluation process should include multiple formats. Participant 2 suggested, "I think there needs to be more than one evaluation [format]. There needs to be alternative ones, too." Participant 6 stated, "If there was one just for beginning teachers, that may be focused more on management . . . I think it would be cool to have different versions." Participant 8 suggested that the NCEES be used for the summative evaluation and a different format be used for the observations.

Two participants, or 8%, suggested that educators need additional training on the NCEES to make it more effective. Participant 4 said, "I don't know what kind of training is actually given to help people understand what each of these [standards] mean." Participant 23 said, "I don't know if our evaluators understand that rubric enough." Participants 16 and 19 did not have any suggestions for improving the evaluation process. Participant 19 does not like the current process, but wasn't sure how it could be improved. Participant 16 was the only participant in the study who likes the system as it is now. Table 24 summarizes these findings.

Participants also provided suggestions on different ways to measure teacher effectiveness in addition to or as opposed to a standards-based observation method and VAD. Participants suggested the following methods: parent input /surveys, student input/ surveys, teacher portfolios and standards-based grading. The most common measurement tool suggested by participants was the parent and student survey. Seventeen percent of participant suggested surveying parents and students. Participant 12 urged, "There needs to be input or insight from both the students and the parents." Participant 18 also suggested input from students and parents. She said, "I think it would be eye opening . . . just to really see how parents and students, the people who are getting our services, feel." Participant 19 also stated, "Feedback from parents and students. Parents are part of the learning community and you can't forget that." Participant 22 suggested, "Possibly do a parent and student survey, maybe take out the best and the worst surveys and kind of get an overall average."

Table 24

Participants' Suggestions on Improving Teacher Evaluation in North Carolina

	Better or Clearer Feedback	Multiple Evaluators	More Training	Devote More Time to a Quality Process	Make Instrument Shorter	Omit VAD from instrument or Improve It	Omit Performance Ratings	Omit Peer Observation	Multiple Formats
P1					X				
P2						X			X
P3						X		X	
P4			X						
P5	X				X	X			
P6							X		X
P7						X			
P8	X	X						X	X
P9					X		X		
P10	X			X	X				
P11				X		X			
P12	X	X						X	
P13	X			X					
P14						X			
P15				X		X		X	
P16									
P17	X			X					
P18		X		X	X				
P19									
P20		X		X				X	
P21	X					X			
P22		X		X			X		
P23			X	X		X			
Totals	**7**	**5**	**2**	**9**	**5**	**9**	**3**	**5**	**3**

One participant (4%) suggested that a portfolio process be used. He stated, "It has to be an ongoing, almost portfolio-type process." One participant (4%) suggested standards-based grading. Participant 5 said, "I would love to see standards-based grading.

We have to hold ourselves accountable and the only way to do it is against the standards. You could look at effectiveness that way in conjunction with the snapshot [observation]."

Summary

Chapter IV reiterated the rationale for the study and proposed the research questions. The study's participants were described and the interview protocol was explained. Results from the collection of data were discussed in detail. Overall, the majority of the participants had negative perceptions about the fairness of the teacher evaluation process and of the NCEES with the greatest number of participants viewing the NCEES as unrealistic for teachers. Almost half of the participants felt that the NCEES is more of a requirement than a tool that actually helps them grow professionally. Other negative perceptions included that teacher evaluation is too subjective, provides limited or no feedback, is too lengthy and/or time consuming, and provides just a snapshot of what teachers do. Participants also felt that many components of the NCEES are not observable and, therefore, should not be included on the instrument or that there should be a form for observations and a form for the summative.

Chapter IV also reported the knowledge that participants had about the standards on the NCEES, how they were rated on the NCEES, and what type of feedback they have received. Very few participants knew the standards on their own and nobody could identify more than one standard. Once participants were reminded of each standard, participants had limited perceptions of what the standards entailed or included. The majority of participants had been rated as proficient and/or accomplished on each standard. Regardless of how participants had been rated, feedback received by

participants was very limited and, therefore, had little to no effect on overall professional practices of participants.

Over half of the participants understood that VAD is based on student growth. Over half of them described their own understanding of VAD as limited. More than half of participants could not explain their own value-added data. Ninety-two percent of participants felt that VAD is unfair and should not be included in the NCEES. The most commonly cited reason for it being unfair to include is that too many factors affect student performance and, thus, affect VAD. The majority of participants could not recall how they were rated on Standard Six of the NCEES, which is the standard based on VAD.

Finally, Chapter IV provided a description of recommendations that participants offered for improving the teacher evaluation process in North Carolina. Over a third of participants suggested evaluators spend more quality time with teachers during the evaluation process. Over a third of participants suggested that VAD be omitted from the NCEES. Almost a third of participants also identified that teachers needed clearer or more specific feedback during the process. Other suggestions included using multiple or outside evaluators, make the NCEES a shorter instrument, omit the peer observation component, using multiple formats of the instrument, omit performance ratings (and just use verbal feedback), and provide more training to teachers and evaluators on the NCEES. Participants also suggested alternative ways to evaluate teacher effectiveness including parent and student surveys, using a portfolio process, or using standards-based grading.

The next chapter will discuss the study's findings as they relate to the conceptual framework introduced in Chapter I. Findings will also be discussed as they relate to possible recommendations for future practice, policymaking, and research. Specifically, Chapter V will discuss ways in which the teacher evaluation process might be improved to increase its impact on the professional practices of teachers and to improve teacher perceptions of the evaluation process. Recommendations will also be discussed related to the utilization of the NCEES and possible revisions of the instrument. Implications for future policy will also be discussed. Finally, the following chapter will provide recommendations for future research on teacher evaluation.

CHAPTER V

CONCLUSION AND RECOMMENDATIONS

Intended Outcomes of the NCEES and Improvements in Utility

The NCEES is supposed to facilitate a process whereby teachers "receive constructive feedback from administrators, peers, and mentors who complete observations" and "engage in critical conversations with evaluators throughout the process and when final ratings have been assigned" (NCDPI, 2012a, p. 1) for the purpose of professional growth (NC State Board of Education Policy Manual, 2011). According to the outcomes of this study, neither of these intended outcomes seems to actually occur in most teachers' experiences. In terms of meaningful and constructive feedback, very few participants reported receiving any. Figure 2 summarizes the nature of feedback received by participants under each standard during the evaluation process. Figure 2 clearly shows that very few teachers are able to recall receiving any constructive feedback. Figure 3 shows the percentage of participants who could provide an example of feedback having an impact on their practices in each standard.

The implementation of the NCEES and the six standards to evaluate teachers, including the use of VAD, was largely facilitated by North Carolina's intentions of winning Race to the Top funding. As Cavanagh (2011) points out, Race to the Top encouraged teacher and union input. Now that NCEES has been implemented for several years, decision makers about the future of teacher evaluation in North Carolina need to

listen to what teachers are telling us about the NCEES. According to the outcome of this study, teachers describe the current instrument as flawed. The working conceptual framework originally described in Figure 1 from Chapter I, therefore, does not accurately reflect the actual evaluation process that occurs in practice.

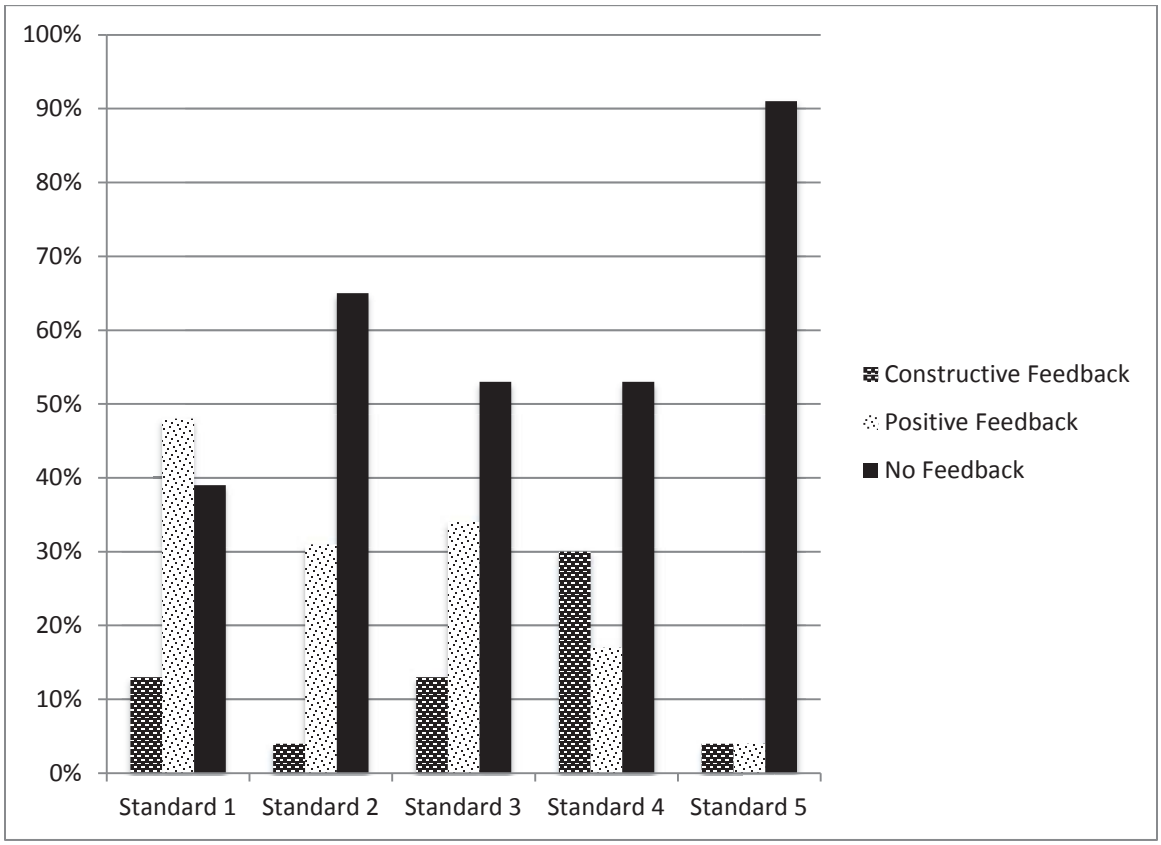

Figure 2. Nature of Feedback during Evaluation Process.

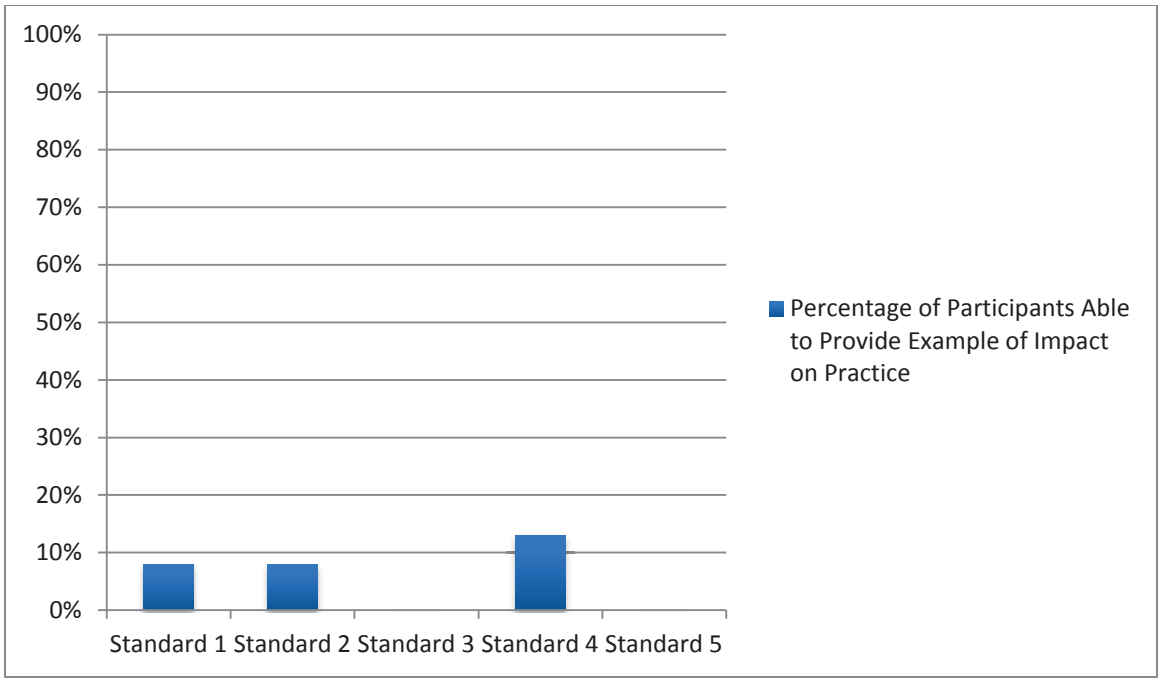

Figure 3. Percentage of Participants Able to Provide Example of Feedback's Impact on Practices

Instead, the evaluation process seems to be little more than a summary of what is happening in the classroom and if feedback is provided, it is usually validating in nature, regardless of the level of performance marked on the evaluation instrument. As one of the participants stated, the evaluation process is "more like a deadline." Another participant recalled, "I can't think of anything, any feedback that's been given to me that would cause me to change practice, which is awful but true." Yet another stated, "I haven't had clear feedback and I haven't had the support to fix it." Kimball (2002) and Feeney (2007) found similar results. Likewise, Heneman III & Milanowski (2003) found that teachers seldom receive in-depth feedback and that the evaluation process holds little value to

them. O'Donovan (2011) argues that teacher evaluation does not make teachers better at teaching and, indeed, this seems to be the case. As O'Donovan (2011) succinctly states,

> Those who think that teacher evaluation . . . can consistently improve teacher practice miss the reality of the process . . . Many teachers put up with the process as a necessary annoyance, but question the validity of the judgment passed by the administrator/evaluator. Minimal observations do not capture the complexity of the teaching and learning processes. Teachers question the validity of the evaluator's comments, as they know that so much of what they do can never show up in an evaluation (p. 73).

Papay (2012) contends that evaluation tools should be judged by their "quality of targeted feedback they provide and by their ability to drive continued instructional improvement" (p. 124). He further argues that improving student learning will occur much more significantly by focusing on the professional growth aspect of teacher evaluation, rather than on the identification of effective or ineffective teaching. Larsen (2005) asserts the same and claims that "placing too much emphasis on the evaluation of individuals for accountability purposes can foster an unproductive climate of fear and detract from the establishment of shared professional responsibility (p. 299). Larsen further contends that federal and state developed "formal assessment procedures [are] a mode of control while appearing to devolve power to individuals and autonomous institutions" (p. 300). Several participant responses reflected this contention. For example, one of the participants likened the evaluation process as "being put in a box." Another participant described the limited amount of time evaluators have to devote to the process and argued, as a result, most teachers get an "average" evaluation rating. She said, "If [the observation] is too good, they have to write. If it's too bad, they have to write, so you just get the middle of

the road no matter what." Likewise, another participant stated, "We perceive it [the evaluation process] as being pressured into conforming to something."

This seems to be where many evaluators are going wrong with the NCEES. Judging by the experiences of participants in this study, most evaluators seem to lack the time or the inclination to providing teachers with meaningful feedback that will improve teacher performance and ultimately improve student outcomes. One of the participants described the process as "a revolving door of due dates." Instead, teachers receive an identification of their performance as developing, proficient, accomplished, or distinguished, without any substantive feedback on how to improve. It does not appear that the NCEES is any different from the many "rubber stamping" standards-based evaluations from years past. If we begin to see the NCEES primarily as a tool for the professional development of teachers, we may truly begin to tap into its potential.

The NCEES may be more effective if an abbreviated version is used during formative evaluations, as it will likely improve perceptions of procedural justice (which is explained in the following section). Kyriakides, Demetriou, and Charalambous (2006) found that the appropriateness of criteria used in teacher evaluations cannot be adequately judged until one determines if each criterion will be used formatively or in a summative manner. These researchers urge that different criteria be used for each type of evaluation. Results from my study support this notion that teachers perceive some criteria as being unfair for classroom observations (the formative component of teacher evaluation). For example, participants suggested that the evaluation process in North Carolina needed "different formats" for classroom observations and to "make it shorter" for classroom

observations. Several participants felt the current NCEES is more appropriate as a summative tool.

The peer evaluation component of the NCEES does not seem to be an effective component. Participants of my study do not report that this component is accurate due to peer's tendencies to avoid giving candid feedback or due to their lack of knowledge about the criterion included in the teacher evaluation instrument as it relates to their peer's performance. For example, one of the participants stated that teachers "don't want to upset their peers" by giving critical feedback. As one of the participants stated, "I don't mind sharing [ideas,] but no one wants to necessarily hear that from someone. I don't want to go telling people how to do the job, necessarily . . . that's offensive." Kyriakides et al. (2006) also found that teachers do not generally perceive peer evaluation positively as an effective form of evaluation. One of the participants in this study said, "[Observing peers] is very uncomfortable . . . you didn't want to focus on the negative because you don't want to have that type of relationship with your coworker."

Improving Teacher Receptivity to the Evaluation Process

As Leary (2007) explains, people have an innate desire to maintain or improve a positive self-concept as a way to protect or enhance their self-esteem. Closely associated to this phenomenon are several tendencies, one of which is to attribute positive events or outcomes to their own doing but attribute negative events or outcomes to factors beyond their control. Another human tendency is to evaluate oneself more positively than other objective individuals would do so. In fact, Leary states that people also "tend to evaluate themselves more positively than they rate the average person on virtually every

dimension that has been studied" (p. 321). In addition, people tend to seek opportunities to validate their existing self-concepts. One way people validate their existing self-concepts is by remembering feedback that verifies their own self-concept and dismissing feedback that is inconsistent with their own self-concept, even when the feedback is accurate (Feys, Libbrecht, Anseel, & Lievens, 2008; Leary, 2007; Sedikides & Green, 2004). This is referred to as mnemic neglect and occurs when feedback is given on central, or more important traits to the individual, as opposed to peripheral, or less important traits to the individual (Sedikides & Green, 2004). Furthermore, as Leary continues, "Ironically, people's tendency to self-enhance also leads them to think they are not self-enhancing" (p. 322).

Human bias toward the self is relevant to performance evaluation processes and related feedback because it is likely a significant factor in the degree of receptiveness to constructive feedback during the evaluation process, to the likelihood that individuals will reflect on ways to improve their own practices and efficacy, and to their interpretations of student outcomes as a variable of their own performance. Feys et al. (2008) maintains that, despite the fact that most organizations evaluate employee performance, dissatisfaction about performance evaluation is not unusual. In fact, Feys et al. (2008) claim that many evaluation systems fail to guide professional development because of two main reasons. First, employees tend not to accept or utilize feedback given when they do not have a positive relationship with the supervisor. Second, employees tend not to accept or utilize feedback given when they perceive that they are being treated unfairly during the evaluation process, which is what Feys et al. (2008)

refer to as procedural justice. As perceptions of procedural justice increase, feedback acceptance also increases. In addition to prerequisites of a positive relationship with the supervisor and perceptions of fairness, the receiver of feedback, the employee, must perceive the feedback as having utility before he or she will demonstrate an openness to utilizing the feedback in an attempt to enhance performance. Both of these factors seem to have such an impact in the arena of teacher evaluation (Heneman III & Milanowski, 2003; Kimball, 2002; O'Donovan, 2011).

My study supports this claim. For example, one of the participants (#9) described the process as "very flawed." This same participant communicated a reluctance to accept feedback throughout the interview. She stated, "Nobody knows if I'm really doing that or not" when discussing Standard Five. Another time, she said, "You learn how to play the game." When asked about using feedback to inform her practices, she said, "Honestly, I haven't put a whole lot of effort into that." Another participant (#12) who communicated negative perceptions about her relationship with the evaluator stated, "I can count on two hands when they come in. I need to feel like when they come in it's not a 'gotcha.' I feel like I'm being targeted." Like Participant 9, she was reluctant to accept feedback. When asked about using feedback to inform her practices, she said, "The evaluation means nothing to me at this point."

Darling-Hammond et al. (2012) urge that InTASC standards be used in conjunction with teacher observations, which is what the NCEES is based upon. It is not enough, however, to have a tool that is well grounded in professional standards and research. Evaluators must receive sufficient training so that the tool will be used with

accuracy and fidelity. Teachers will be more receptive to the evaluation process if they are confident that the evaluator has been properly trained. Evaluators also must devote the proper amount of time to the process, spending sufficient time in the classroom outside of the evaluation process and during the evaluation process so that teachers feel they have a sufficient level of awareness about their daily practices. Also, feedback needs to be specific and given immediately after the observation. As Darling-Hammond et al. (2012) state, "Successful systems use multiple classroom observations across the year by expert evaluators looking at multiple sources of data, and they provide timely and meaningful feedback to the teacher" (p. 13). Clearly, that is not happening for some of the participants in this study. As one participant claimed, "I was evaluated a few weeks ago and I still haven't been called to sign off on anything or to discuss it."

Bill and Melinda Gates Foundation has developed measures of teacher effectiveness through their Measures of Effective Teaching (MET) Project that may serve to assess teachers in a fair manner. They suggest that teachers be observed multiple time per year by more than one evaluator, using an observation rubric, and using student surveys in addition to student achievement data (Marshall, 2012). However, they do not recommend an elaborate rubric, as it "creates an impossible workload for administrators, leaving less time for informal classroom visits and interactions with teacher teams" (Marshall, 2012, p. 50). Kyriakides et al. (2006) describe the use of the school constituencies satisfaction model, whereby teacher quality is assessed, in part, through student surveys (like the MET Project suggests) but also with parent surveys. Several participants suggested that the teacher evaluation process in North Carolina include

student and parent surveys. As one of the participants argued, “Students and parents are part of the learning community and you can’t forget that.” Another participant said it would be “eye opening” to get feedback from parents and students. Numerous other participants voiced similar support for student and parent surveys.

Danielson and McGreal (2000) recommend four primary aspects in a standards-based evaluation tool: 1) a detailed rubric of teacher performance, 2) multiple levels of performance clearly outlined, 3) frequent observations and multiple sources of evidences and artifacts, and 4) evaluators that are well trained. It seems that the first two criteria are met by the NCEES, but the last two are questionable. As stated previously, numerous participants claim that observations are few and often rushed and some participants voiced concern over the lack of consistency and training of evaluators. As one participant voiced, “I don’t know what kind of training is actually given to help people understand what each of these [elements] mean.” This same participant (#4) described her experiences with being evaluated that varied drastically from evaluator to evaluator, despite her contention that her practices remained the same. This lack of training for evaluators, she felt, was a possible explanation for this.

Recommendations for Changes in Policy and Practice

There is an important relationship between policy adoption and policy implementation and the monitoring of policy implementation (Christie & Fierro, 2012). Teaching and teacher evaluation practices rest on a multitude of educational policies that have been driven by public demands for accountability (Larsen, 2005). A breakdown in

this relationship can result in practices that are not consistent with policy and/or policies that are meaningless in day-to-day practice.

Christie and Fierro (2012) argue that insufficient research has been conducted to determine how teacher evaluation policies are being interpreted and implemented by evaluators and, thus, how they are impacting the teachers who are affected by such policies. These researchers further contend that educational policies are often written, but "are not always carried out as intended" (p. 66). One reason they cite for this discrepancy is the lack of trained evaluators, who may possess content area expertise but do not have expertise in evaluative and experimental methods, which are supposed to be based on objective practices. The typical evaluator, however, translates policy "through a lifetime of experience, perceptions, and values" that may not be consistent with the policy's original intentions (p. 71). Larsen (2005) argues that education has been vulnerable to the application of external, business management theory in public education because education management has not had its own theory on which to base management practices. In addition, she cites that all sectors are vulnerable to the pressures of "market-driven global forces" which may include employee quality improvement, modernization, and concepts of accountability (p. 293).

In a survey conducted with principals, Johnson (2004) found that only 11% of principals believe that teacher certification guarantees that teachers have what it takes to be effective. If this is accurate, it would suggest that solving issues related to teacher quality need to focus on improving teacher preparation programs, since this statistic indicates that preparation programs are vastly inadequate in preparing teachers to be

effective. Guthrie (2009) elaborates by describing the traditional student teaching as "seldom rigorous or comprehensive. It usually results in a simple letter grade" (p. 9). He contends that prospective teachers should be evaluated in terms of basic knowledge and communication skills, expertise in subject matter, reading and math instructional skills, ability to differentiate, classroom management, capacity to build rapport with a diversified group of students, ability to construct and interpret tests, and communication skills with parents and other community members. By increasing the rigor of how we evaluate the readiness of these prospective teachers, he argues that we can more accurately identify who is adequately prepared to enter the teaching profession as a capable and effective teacher.

Waddell (2012) agrees. He argues that many programs are little more than "diploma mills" (p. 26). He also argues that the Educational Testing Service allows teacher candidates to "perpetually retest" without any delay, which can allow some candidates to score barely well enough to pass (p. 26). Perhaps standards should be raised on such tests, if candidates scoring at the minimum passing rate are demonstrating a lack of proficiency in the field.

According to Mourshed and Barber (2007), the world's best performing school systems recruit teachers from the top third of their graduate cohort, while the United States (with the exception of specific teaching programs such as Teach for America, Boston Teacher Residency, and both the New York and the Chicago Teaching Fellows programs) is recruiting teachers "from the bottom third of high school students going to college" (p. 16). Thomas et al. (2010) cite that as much as a third of teachers have

historically scored in the lowest 5% on the SAT. Mourshed and Barber (2007) argue that it is not possible to increase the quality of education without improving the quality of our teachers. They point out that top performers screen potential teaching candidates before entering teacher training programs, rather than screening candidates right before hiring, which is what low performing systems do, like the U.S. Darling-Hammond (2007) agrees that we must be willing to invest more, like other better performing countries, in preparing teachers than we currently do. In addition, we must be willing to provide our more disadvantaged students from lower socioeconomic status access to high-quality teachers (Murnane & Steele, 2007). To do this, it would require a major transformational shift in our commitment to high-quality teacher preparation programs, a strong investment in paying our teaching force, and a willingness to provide all schools with equitable resources, regardless of their location or who they serve.

According to the National Council on Teacher Quality (NCTQ), the state of North Carolina earned a D in the area of adequately preparing teachers for the profession on the 2011 State Teacher Policy Yearbook and on the 2012 State Teacher Policy Yearbook. The NCTQ identified 76% of teacher preparation programs as being inadequately selective in choosing teacher candidates. Only the state of Texas requires a test of academic proficiency that is normed to the general college population, as opposed to being normed against only prospective teachers. In 2013, the NCTQ gave North Carolina a C for preparing new teachers, which was an improvement, though we continue not to require a minimum grade point average (GPA) for admission to many teacher preparation programs. A likely reason for the improvement in NCTQ's rating is probably because

North Carolina is finally requiring elementary teacher candidates to pass a Foundations of Reading test. However, NCTQ stated in the 2013 State Teacher Policy Yearbook, "North Carolina should evaluate its passing score to make certain it reflects a high standard of performance" (p. 18).

Considering the statistics for those entering the teaching profession, this standard is not high enough if we want to improve the quality of teaching candidates entering the profession. In addition, North Carolina does not require elementary teachers to be prepared or knowledgeable in "the science of reading" instruction; their knowledge in reading instruction is not measured. This is another major flaw with North Carolina's teacher preparatory programs according to the National Council on Teacher Quality. North Carolina is one of only a few states for which the passing score on content licensure tests is not available for public review, though over 30 states set their passing score as at least one standard deviation below the mean. It is not surprising, therefore, that low standards for entry into the teaching profession ultimately means poor outcomes for many students.

Indeed, since 1983's *A Nation at Risk*, little has changed. According to the U.S. Department of Education's *A Nation Accountable*, published 25 years after *A Nation at Risk*, students' reading and math proficiency levels have remained stagnant, students spend no more time in school than they did in 1983, and only 53% of African Americans and 58% of Hispanic students in urban schools graduate on time.

Teachers need to receive more feedback and better feedback (Feeney, 2007). In Harrison School District, Colorado, teachers are observed up to 16 times each semester.

Each time they receive feedback (Goldstein, 2011). Evaluators must be willing and able to provide specific and critical feedback, but many of them are either not willing or able (Morgan, 2012). The results of this study found that most teachers do not receive adequate feedback that would lead to any kind of improvement in practices. Part of the reason for this phenomenon seems to be related to the fact that many administrators are not willing to initiate the levels of support and documentation required prior to removing ineffective teachers (Morgan, 2012). As Waddell (2012) explains, "evaluations are often superficial and conducted by harried administrators" (p. 26). Kimball (2002) had such findings in his research. Another reason is that some administrators are not comfortable with giving candid feedback (Johnson, Fiarman, Munger, Papay, & Qazilbash, 2009).

The resulting outcome is what Weisberg, Sexton, Mulhern, and Keeling (2009) refer to as the Widget Effect. This is the phenomenon that describes the tendency for evaluations to be similar from teacher to teacher, regardless of the actual differences in effectiveness. The phenomenon is characterized by five actualities: 1) all teachers are rated as good or great, 2) excellence is not recognized or rewarded, 3) professional development is inadequate and not individualized, 4) novice teachers do not obtain adequate support or attention, and 5) poor performance is typically not addressed.

To reverse the Widget Effect, researchers recommend that administrators must receive adequate training on the evaluation tool being utilized and the evaluation process must be implemented with fidelity (Heneman III & Milanowski, 2003; Kimball, 2002; Papay, 2012; Weisberg et al., 2009). Papay (2012) describes the evaluation process in Cincinnati, where evaluators must complete an extensive training that includes a video

component, on which evaluators "need to achieve sufficient agreement on these lessons in order to be certified" (p. 130). Even after certification, evaluators meet twice each month for additional training that includes calibration of ratings. Such training could also address and improve administrators' confidence and comfort level with having candid conversations with teachers, which some evaluators struggle with (Johnson et al., 2009).

Effective feedback is characterized by its descriptiveness, inclusion of effective teaching characteristics, and its ability to promote reflective inquiry and improvements in teaching (Feeney, 2007; Looney, 2011). As the results of this study show, most teachers do not experience this type of feedback during the evaluation process. Therefore, evaluator training must include a component whereby they practice generating effective feedback in addition to calibrating level of performance ratings.

Teachers also need to be given adequate time to plan with one another and observe each other teaching for the purpose of professional growth, similar to countries such as Finland and South Korea, who outperform the United States on international tests of achievement (Goldstein, 2011). Collaboration and professional inquiry focused on analysis of student learning and improved student performance has greater power to improve teacher performance (Looney, 2011; O'Donovan, 2011). As Larsen (2005) states, "we need to focus on creating systems and environments that foster excellence and recognize that teaching is work that is creative, continually changing, pluralistic, diverse, and complex" (p. 302).

The Use of VAD in Teacher Evaluation

Standardized tests do not seem to be strong enough measures to be used for high-stakes purposes. The difference between two scores on a standardized test must be equivalent to the same difference between two other sets of scores on a standardized test and this is not the case (Doran & Fleischman, 2005; Galley, 2011; Gitomer, 2011). Additionally, VAD seems to have limited consistency from year to year. VAD is effective with identifying the very best and the very worst of teachers, but it has limited utility for most teachers who perform in the middle of the top and bottom quintiles (Corcoran & Goldhaber, 2013). In addition, VAD does not make teachers better at what they do as it gives no information about what the teacher actually did and teachers seem not to be able to gain much reflection about their practices from their own VAD (Darling-Hammond et al., 2012; Jones & Egley, 2006). My study found similar results. Larsen (2005) argues that "the attempt to capture the complexity of teaching in paper-and-pencil tests and performance assessment tools has resulted in the simplification of the art of teaching into linear testing formats and performance competency checklists" (p. 298). The fact that VAD simplifies something that is very multifaceted and complex is a strong point.

Other sources besides value-added data could be considered for Standard Six of the NCEES. For example, in Tennessee, only 35% of the evaluation standard related to student outcomes uses VAD. Fifteen percent uses sources agreed upon by the teacher and evaluator (Green et al., 2012). Arizona uses a system whereby teachers use pre- and post-tests that are created at the classroom level or school level to show evidence of student

learning. Such a practice has the benefit of having better buy-in from teachers, but also teachers tend to improve their skills at creating assessment tools and show heightened awareness of and focus on better quality instruction (Darling-Hammond et al., 2012).

Summary

Barile (2013) argues that teacher accountability "has become the predominant rallying cry of politicians and education critics alike, who have singled teachers out as scapegoats for all of the failings in American education today" (p. 127). It is true that public education has a multitude of issues, such as limited funding and resources, overcrowded schools, outdated facilities, and so on. Good et al. (2010) also argues that state and local funding policies also contribute to these issues, since it leads to inequitable funding for schools. The result is that different teachers work with different access to resources. Good et al. (2010) assert,

> Most teachers are working hard and often under very difficult conditions. Teachers are expected to work hard and wisely, but they alone cannot offset the terrible toll that poverty exerts on learning. Those who support the current high-stakes testing movement may not realize how this movement, whether intentional or not, has devalued teachers. (p. 146)

Nevertheless, those issues cannot prevent us as educators from holding ourselves accountable for student outcomes. Accountability, however, must be fair. In many professions, outcomes are used to measure performance. However, as one of the participants in this study concisely stated, "I built houses for . . . years. You're evaluated by your customers, but you had control on your parts and pieces, your inventory, and the things that you use in production. In education, you don't have control." The participant

was referring to limiting factors such as pacing guides, textbook adoptions, schedules, and even expectations from administrators surrounding lesson formatting and instructional strategies. Papay (2012) contends,

> Evaluations must provide teachers with a clear understanding not only of their current success or failure, but also of the practices they need to develop to become more successful with their students. There are few examples of current systems that effectively combine these two purposes, effectively measuring teacher performance and providing feedback to help them improve. (p. 138)

An effective system of teacher evaluation needs to use multiple sources of data, because "no single measurement can capture the full range of teacher performance in different contexts or conditions, or the qualities that are important for effective teaching" (Looney, 2011, p. 443). Even proponents of VAD urge the use of multiple data sources to evaluate teachers (Corcoran & Goldhaber, 2013; Hanushek & Rivkin, 2010). Not only do multiple sources increase accuracy, they also provide different types of information for the evaluator that can provide a more holistic view of the teacher as professional educator (Looney, 2011; Richardson, 2012). Multiple measures may include observations, student outcomes as measured by multiple sources, and parent and student surveys (Richardson, 2012).

Using the current review of research and the current findings of this study, a revised contextual framework is offered using the input of Stronge et al. (2011) and Kimball (2002). This revised contextual framework is illustrated in Figure 4. As the revised contextual framework illustrates, the evaluator looks for the various qualities of effective teaching during standards-based evaluation.

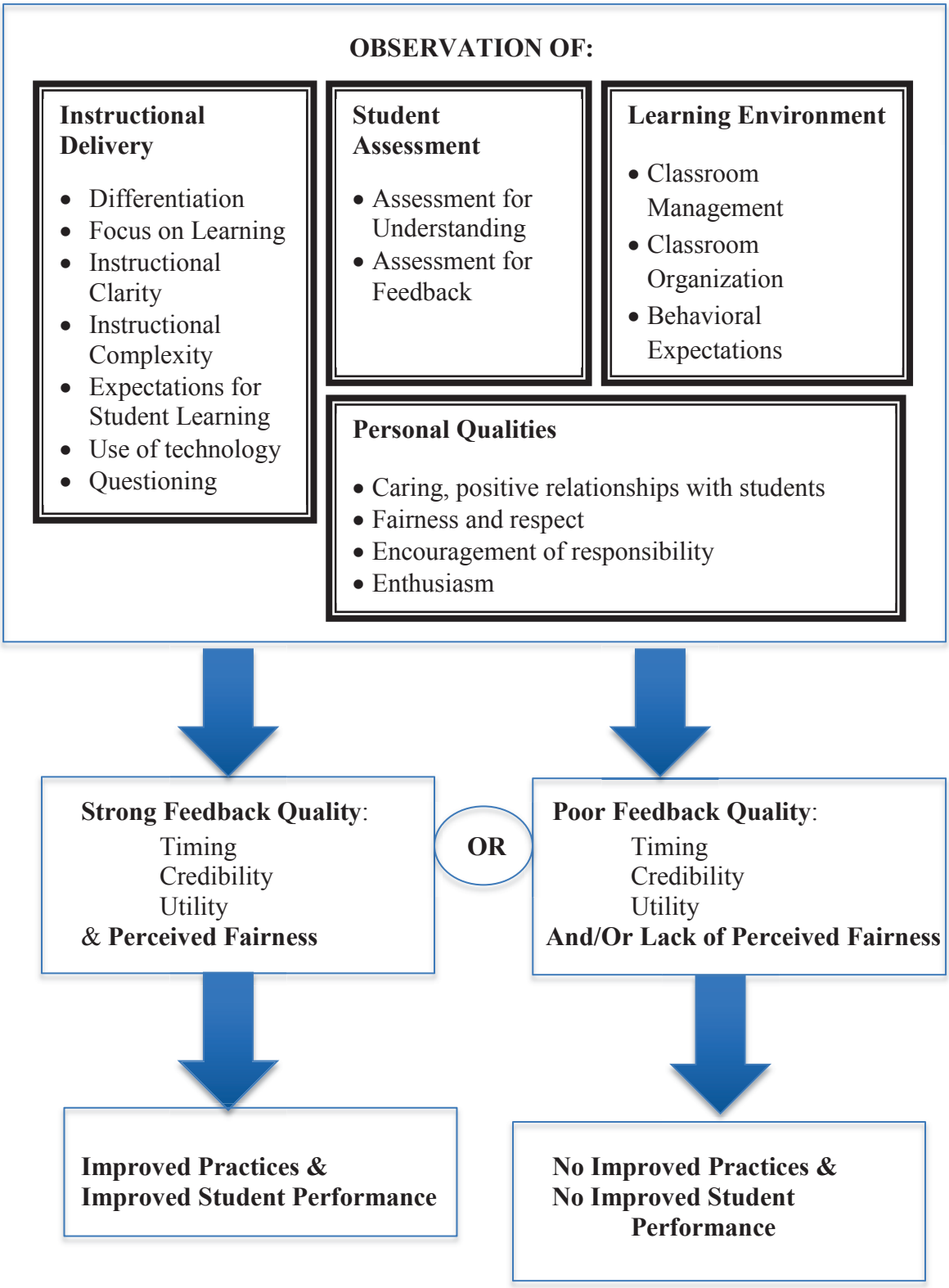

Figure 4. Revised Contextual Framework.

Subsequently, the evaluator provides feedback of strong quality or poor quality. Strong feedback quality is characterized by timeliness, utility, and credibility. As Kimball (2002) explains, effective feedback is given immediately, it is specific and contains accurate information related to the content or area observed, and it is useful to the recipient. Poor feedback is lacking in one or more of these areas.

The other important component during the standards-based evaluation process is the perception of fairness by the person being evaluated. If the individual perceives the process to be fair (procedural justice), then strong quality feedback can be received and internalized. If the individual does not perceive the process to be fair (a lack of procedural justice), then the feedback, regardless of quality, will not be internalized. These factors determine if the evaluation process will ultimately result in any sort of improved practice and, therefore, improvements in student outcomes.

Teacher quality and the professional growth of teachers are perpetual domains of focus for an instructional leader, since we understand that the quality of a child's education is greatly dependent upon the quality of instruction in the classroom. Prior to this study, I was confident as an instructional leader and evaluator of teachers that my efforts to fairly evaluate teachers and to provide quality feedback were worthwhile endeavors. Of course teachers deserve quality feedback about their performance, but I feel that my perspective has been broadened and enlightened by the realities of teacher experiences with the NCEES and with the evaluation process that were shared with me during this study. Despite the comprehensiveness of the NCEES, the participants of the study have highlighted the limited outcomes of the tool and of the process. Ironically, it

seems the comprehensiveness of the tool hinders its utility and, in actual practice, is excessive, burdensome, and has limited surface validity with those who are evaluated. The result of this discord between theory and practice is an inordinate amount of wasted time for both teachers and leaders, because the practice of evaluation with the NCEES is neither ensuring teacher quality (or differentiating levels of quality) nor promoting improvement in professional practices.

Ultimately, in order to improve the quality of education in our public schools, it seems that we need to make a significant shift from placing blame among our teachers to establishing programs in our institutions of learning that are rigorous and then providing teachers who enter the workforce with an effective evaluation process that offers quality feedback and opportunities for growth. We frequently ask teachers to reflect upon their data. When will we begin to reflect upon ours?

REFERENCES

American Recovery and Reinvestment Act. (2009). Public Law 111-5.

Barile, F. (2013). Making enemies out of educators: The legal and social consequences of disclosing New York City teacher data reports. *Brigham Young University Education and Law Journal, 1,* 125–142.

Cavanagh, S. (2011). Courthouse rife with education policy fights. *Education Week, 31*(5), 1–19.

Chapman, C., Laird, J., Ifill, N., & KewalRamani, A. (2011). Trends in High School Dropout and Completion Rates in the United States: 1972–2009 (NCES 2012-006). *U.S. Department of Education.* Washington, DC: National Center for Education Statistics. Retrieved November 11, 2014 from http://nces.ed.gov/pubsearch

Chelsey, G. M. (2011). Now is the time to redefine teacher tenure. *Education Week, 30*(29), 34–35.

Christie, C. A., & Fierro, L. A. (2012). Evaluation policy to implementation: An evaluation of scientifically based research in practice. *Studies in Educational Evaluation, 38*(2), 65–72.

Coble, R. (2000). North Carolina: First in education reform? *North Carolina Insight*, 66–67. Retrieved from www.nccppr.org

Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington, DC: National Center for Educational Statistics.

Corcoran, S. P., & Goldhaber, D. (2013). Value-added and its uses: Where you stand depends on where you sit. *Education Finance and Policy, 8*(3), 418–434.

Corcoran, S. P., Jennings, J. L., & Beveridge, A. A. (2011). *Teacher effectiveness on high- and low-stakes tests.* Retrieved from http://files.nyu.edu/scl29/public/papers/corcoran_jennings_beveridge_2011_wkg_teacher_effects.pdf

Danielson, C., & McGreal, T. L. (2000). *Teacher Evaluation to Enhance Professional Practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Darden, E. C. (2013). Firing a teacher is getting easier. *Phi Delta Kappan, 94*(4), 68–69.

Darling-Hammond, L. (2007). The flat earth and education: How America's commitment to equity will determine our future. *Educational Researcher, 36*(6), 318–334.

Darling-Hammond, L. (2008). Educating teachers: How they do it abroad. *Time, 171*(8), 34.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8–15.

DeMitchell, T. A., DeMitchell, T. A., & Gagnon, D. (2012). Teacher effectiveness and value-added modeling: Building a pathway to educational malpractice? *Brigham Young University Education and Law, 2,* 257–301.

Di Carlo, M. (2012). How to use value-added measures right. *Educational Leadership, 70*(3), 38–42.

Doran, H. C., & Fleischman, S. (2005). Challenges of value-added assessment. *Educational Leadership, 63*(3), 85–87.

Duffrin, E. (2011). What's the value in value-added? *Education Digest, 77*(2), 46–49.

Feeney, E. J. (2007). Quality feedback: The essential ingredient for teacher success. *Clearing House, 80*(4), 191–198.

Feys, M., Libbrecht, N., Anseel, F., & Lievens, F. (2008). A closer look at the relationship between justice perceptions and feedback reactions: The role of the quality of the relationship with the supervisor. *Psychologica Belgica, 48*(2–3), 127–156.

Frank Cona v. Avondale School District, 303 Mich App 123 (2014).

Galley, L. A. (2011). The research vs. the rhetoric. *Education Digest, 77*(3), 41–46.

Gitomer, D. H. (2011). Road maps for learning and teacher evaluation. *Measurement, 9*, 146–148.

Goldhaber, D., & Hansen, M. (2010). *Is it just a bad class? Assessing the stability of measured teacher performance.* Center for Education Data and Research. Seattle, WA: University of Washington.

Goldstein, D. (2011). The test generation. *American Prospect, 22*(4), 14–19.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context (NCES 2009–001 Revised). *U.S. Department of Education*. Washington, DC: National Center for Education Statistics.

Good, T. L., Wiley, C. R., & Sabers, D. (2010). Accountability and educational reform: A critical analysis of four perspectives and considerations for enhancing reform efforts. *Educational Psychologist, 45*(2), 138–148.

Gordan, R., Kane, T. J., & Staiger, D. O. (2006). Identifying effective teachers using performance on the job. *The Hamilton Project.* Washington, DC: The Brookings Institution.

Green, J. D., Sedikides, C., Pinter, B., & Van Tongeren, D. R. (2009). Two sides to self-protection: Self-improvement strivings and feedback from close relationships eliminate mnemic neglect. *Self and Identity, 8*(2–3), 233–250.

Green III, P. C., Baker, B. D., & Oluwole, J. (2012). The legal and policy implications of value-added teacher assessment policies. *Brigham Young University Education and Law Journal, 1,* 1–29.

Guthrie, J. W. (2009). An "Education Professions Performance Development Act": A prospectus for providing "highly qualified" and more motivated teachers and leaders for America's schools. *Peabody Journal of Education, 80*(3), 6–14.

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

Hanushek, E. A. (2007). The single salary schedule and other issues of teacher pay. *Peabody Journal of Education, 82*(4), 574–586.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review, 30*(3), 466–479.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2). Amsterdam: North-Holland.

Hanushek, E., A. & Rivkin, S. (2010). Generalizations about using value-added measures of teacher quality. *American Economics Review, 100*(2), 267–271.

Heneman III, H. G., & Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(2), 173–195.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children, 17*(1), 129–153.

Jennings, J. L., & Corcoran, S. P. (2009). Beware of geeks bearing formulas. *Phi Delta Kappan, 90*(9), 635–639.

Johnson, J. (2004). What school leaders want. *Educational Leadership, 61*(7), 24–27.

Johnson, S. M., Fiarman, S. E., Munger, M. S., Papay, J. P., & Qazilbash, E. K. (2009). *A user's guide to peer assistance and review.* Retrieved from http://www.gse.harvard.edu/~ngt/par

Jones, B. D., & Egley, R. J. (2006). Looking through different lenses: Teachers' and administrators' views of accountability. *Phi Delta Kappan, 87*(10), 767–771.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practice using student achievement data. *Journal of Human Resources, 46*(3), 587–613.

Kelly, D., Xie, H., Nord, C.W., Jenkins, F., Chan, J.Y., and Kastberg, D. (2013). *Performance of U.S. 15-Year-Old Students in Mathematics, Science, and Reading Literacy in an International Context: First Look at PISA 2012* (NCES 2014-024). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Kimball, S. M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education, 16*(4), 241–268.

Klein, J. (2011). Scenes from the class struggle. *Atlantic Monthly, 307*(5), 66–77.

Kristof, N. D. (2011). Pay teachers more. *The New York Times.* Retrieved October 12, 2013 from http://www.nytimes.com/2011/03/13/opinion/13kristof.html

Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research, 48*(1), 1–20.

Larsen, M. A. (2005). A critical analysis of teacher evaluation policy trends. *Australian Journal of Education, 49*(3), 292–305.

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology, 58,* 317–344.

Lewis, L., Parsad, B., Carey, N., Bartfai, N., Farris, E., & Smerdon, B. (1999). Teacher quality: A report on the preparation and qualifications of public school teachers. *Education Statistics Quarterly, 1*(1),

Lichtman, M. (2010). *Qualitative Research in Education: A User's Guide.* Washington, DC: Sage.

Looney, J. (2011). Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education, 46*(4), 440–455.

Marshall, K. (2012). Fine-tuning teacher evaluation. *Educational Leadership, 70*(3), 50–53.

Mathews v. Eldridge, 424 U.S. 319 (1976).

Morgan, H. (2012). Poverty-stricken schools: What we can learn from the rest of the world and from successful schools in economically disadvantaged areas in the U.S. *Education, 133*(2), 291–297.

Mourshed, M., & Barber, M. (2007). *How the world's best performing school systems come out on top*. Washington, DC: McKinsey and Company.

Murnane, R. J., & Steele, J. L. (2007). What is the problem? The challenge of providing effective teachers for all children. *Future of Children, 17*(1), 15–43.

National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: U.S. Department of Education.

National Council on Teacher Quality. (2011). *State Teacher Policy Yearbook*. Washington, DC: National Council on Teacher Quality.

National Council on Teacher Quality. (2012). *State Teacher Policy Yearbook*. Washington, DC: National Council on Teacher Quality.

National Council on Teacher Quality. (2013). *State Teacher Policy Yearbook*. Washington, DC: National Council on Teacher Quality.

No Child Left Behind Act. (2001). 20 U.S.C. 6301.

North Carolina Department of Public Instruction. (2012a). *Measuring growth for educator effectiveness: A guide to the use of student growth data in the evaluation of North Carolina teachers.* Raleigh, NC. Retrieved from http://www.ncpublicschools.org/docs/effectiveness- model/ncees/measure-growth-guide.pdf

North Carolina Department of Public Instruction. (2012b). *North Carolina teacher evaluation process*. Raleigh, NC. Retrieved November 15, 2014 from http://www.ncpublicschools.org/effectiveness-model/ncees

North Carolina Department of Public Instruction. (2013). *Free and reduced student data by site*. Retrieved from http://www.ncpublicschools.org/fbs/resources/data/

North Carolina State Board of Education Policy Manual. (2011). Policy ID Number TCP-C-004. Retrieved November 10, 2013 from http://sbepolicy.dpi.state.nc.us/policies/TCP-C-004.asp?pri=02&cat=C&pol=004&acr=TCP

O'Donovan, E. (2011). Teacher evaluation is not synonymous with teacher quality. *District Administration, 47*(7), 73–74.

Olson, A. M., & Sabers, D. (2008). Standardized tests. In T. L. Good (Ed.), *Twenty-first century education: A reference handbook.* Thousand Oaks, CA: Sage.

Paige, M. (2013). Applying the paradox theory: A law and policy analysis of collective bargaining rights and teacher evaluation reform from selected states. *Brigham Young University Education and Law Journal, 1,* 21–43.

Paine, S. L., & Schleicher, A. (2011). *What the U.S. can learn from the world's most successful education reform efforts*. McGraw-Hill Research Foundation. Retrieved from http://www.mcgraw-hillresearchfoundation.org/

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*(1), 123–141.

Pascopella, A. (2006). Rotten apples. *District Administration, 42*(5), 66–74.

Perry, I. (1977). Suppose we lost tenure? *Educational Leadership, 35*(3), 183–186.

Peter W. v. San Francisco Unified School District, 60 Cal. App. 3d 815 (1976).

Race to the Top Application (2010). *Office of the Governor: State of North Carolina.* Retrieved from http://www2.ed.gov/programs/racetothetop/phase2-applications/north-carolina.pdf

Richardson, J. (2012). Million-dollar baby. *Phi Delta Kappan, 94*(3), 4.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252.

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), 247–256.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*(3), 299–311.

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

Sawchuck, S. (2012). Access to teacher evaluations divides advocates. *Education Week, 31*(26), 1–18.

Schachter, R. (2010). A new dawn in the sunshine state. *District Administration, 46*(6), 41–46.

Sedikides, C., & Green, J. D. (2004). What I don't recall can't hurt me: Information negativity versus information inconsistency as determinants of memorial self-defense. *Social Cognition, 22*(1), 4–29.

Shakman, K., Riordan, J., Sanchez, M. T., DeMeo Cook, K., Fournier, R., and Brett, J. (2012). *An examination of performance-based teacher evaluation systems in five states*. (Issues & Answers Report, REL 2012-No. 129). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved November 10, 2013 from http://files.eric.ed.gov/fulltext/ED529800.pdf

Spring, J. (2011). *The American school: A global context from the Puritans to the Obama Era* (8th ed.). New York, NY: McGraw Hill.

St. Louis Teachers Union, Local 420 v. Board of Education of St. Louis, 652 F. Supp. 425 (E.D. Missouri 1987).

State Board of Education (2001). *History of SBE.* Raleigh, NC: State Board of Education. Retrieved from http://stateboard.ncpublicschools.gov/about-sbe/history

Stover, D. (2011). Under fire. *The American School Board Journal, 198*(4), 18–22.

Stronge, J. H. (2002). *Qualities of effective teachers.* Alexandria, VA: Association of Supervision and Curriculum Development.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education, 62*(4), 339–355.

Thomas, E., & Wingert, P. (2010). Schoolyard brawl. *Newsweek, 155*(11), 30–32.

Thomas, E., Wingert, P., Conant, E., & Register, S. (2010). Why we can't get rid of failing teachers. *Newsweek, 155*(11), 24–27.

Toch, T. (2010/20111). Teacher unions are dead! Long live teacher unions! *Phi Delta Kappan, 92*(4), 72–73.

U. S. Department of Education (2008). *A Nation Accountable: Twenty-Five Years after A Nation At Risk*. Washington, DC: U. S. Department of Education.

U. S. Department of Education (2009). *Race to the Top Executive Summary*. Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/programs/racetothetop/index.html

Waddell, C. (2012). The tenure debate. *The American School Board Journal, 199*(10), 25–27.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect. *Education Digest, 75*(2), 31–35.

Winters, M. A. (2012). How to grade teachers. *National Affairs, 13*, 61–76.

APPENDIX A

ORAL RECRUITMENT SCRIPT

I am a doctoral student at the University of North Carolina-Greensboro. The research study that I have chosen for my dissertation is about teacher perceptions and understanding of the North Carolina Educator Evaluation System, including its use of value-added data as a measure of accountability. The purpose of this study is to explore teacher perceptions of performance accountability measures within the context of North Carolina's new teacher evaluation system. Specifically, this study seeks to examine the understanding and perceptions teachers have of the North Carolina Educators Evaluation System (NCEES), including the recent addition of value-added data and Standard Six of the NCEES. This study also seeks to examine how teachers' experiences with the NCEES affect their classroom practices. Finally, this study will explore ideas teachers have as alternative ways to evaluate teacher performance. Your input is valuable about this process in North Carolina and I am very interested in hearing your thoughts and experiences related to the evaluation system.

I am asking teachers in grades 4-8 who teach English Language Arts or Math to participate in the study. No other criteria are being used for selecting participants for the study. If you agree to participate in this study, you will be asked to participate in an interview of approximately 90 minutes to 2 hours in length and/or participate in a follow-up interview. The interview will be conducted at a time and location of your choice and convenience.

Your identity, school, or district in which you work will not be used. Therefore, your responses in the interview will be confidential. Your identity will be confidential and your responses will not be shared with your principal or anybody else. I will need to audio-tape

interviews so that I may transcribe the interview at a later time. You will be allowed to view the transcription if you would like to do so to check for accuracy.

If you would be interested in helping me examine teachers' perceptions and understanding of the NCEES, please record your name and contact information on the cards provided and I will contact you to schedule an interview.

APPENDIX B

LETTER OF INVITATION TO PARTICIPANTS AND CONSENT FORM

Dear Participant,

You are invited to participate in a research study about teacher perceptions of the North Carolina Educator Evaluation System (NCEES). The identity of each participant will remain anonymous. Any new information that comes up during the study will be provided to you if the information might affect your willingness to continue participation in the project.

You will be asked to participate in an interview of approximately 90 minutes to two hours in length or you may be asked to participate in a follow-up interview. You will be asked questions about your perceptions and understanding of the NCEES and about your personal experiences with the NCEES. You may be asked to review researchers' analysis of your responses to check for accuracy. You will remain anonymous in this study; your name will not be used, nor will the school's name be used to which you are affiliated. You may feel some feelings of discomfort or hesitancy in sharing personal thoughts, feelings, and experiences related to being evaluated on the NCEES. Your responses will in no way have any bearing upon your role at the school or your relationship with the researcher. However, if you have any questions, please contact Dr. Carl Lashley at 336-256-0156 or April Conley at 336-337-4709.

You are being asked to participate in this study because you are a teacher working in grades 4 through 8 at one of the schools chosen to participate in this study and because you are a teacher evaluated on the NCEES. Neither gender nor ethnicity is a criterion for selection.

A possible benefit of the study is that you will feel a sense of satisfaction in sharing any concerns or suggestions you have related to the teacher evaluation process in North Carolina. Researchers hope to identify teachers' perceptions, attitudes, and understanding of the NCEES so that decision makers can consider teacher input in moving forward with the teacher evaluation process in North Carolina.

There are no costs to you or payments made to you for participating in this study.

April Conley has answered all of your current questions about you being in this study. Any other questions concerns or complaints about this project or benefits or risks associated with being in this study can be answered by Dr. Carl Lashley who may be contacted at (336) 256-0156 or at carl.lashley@gmail.com

You are free to refuse to participate or to withdraw your consent to be in this study at any time. There will be no penalty or unfair treatment if you choose not to be in the study. Being in this study is completely voluntary.

Your privacy will be protected. You will not be identified by name or other identifiable information as being part of this project. Audio-recordings of the interviews, transcripts from audio-recordings, and copies of signed consent forms will be kept by the student researcher, April Conley, in a locked container. Electronic copies of interview transcripts will be kept on the student researcher's computer and they will be protected by password.

By signing this form, you are agreeing that you are 18 years of age or older. You also agree to participate in the study described to you by April Conley.

_____ _____

Participant's Signature Date

Contact Information:
Please include a phone number or email address below so that you may be contacted to schedule the interview.

Name (Please Print):
_____

Email:
_____

Phone:
_____

APPENDIX C

INTERVIEW PROTOCOL

General

1. How many years of experience do you have as a teacher, not including this year?
2. Do you currently have tenure status in North Carolina?
3. Please summarize your educational background.
4. What grade(s) and subject(s) do you currently teach?

Knowledge of Standards-Based Teacher Evaluation

5. What do you think about the NCEES? (This is a broad question that will hopefully elicit a rich and detailed response.)
6. What elements of performance appear in Standard One?
7. Tell me about your understanding of and perceptions about Standard One of the NCEES: Teachers demonstrate leadership.
8. How have you been rated under Standard One and what type of feedback have you received?
9. Give some examples of how you have used feedback in this area to inform your practices.
10. What elements of performance appear in Standard Two?
11. Tell me about your understanding of and perceptions about Standard Two: Teachers establish a respectful environment for a diverse population of students.
12. How have you been rated under Standard Two and what type of feedback have you received?

13. Give some examples of how you have used feedback in this area to inform your practices.
14. What elements of performance appear in Standard Three?
15. Tell me about your understanding of and perceptions about Standard Three: Teachers know the content they teach.
16. How have you been rated under Standard Three and what type of feedback have you received?
17. Give some examples of how you have used feedback in this area to inform your practices.
18. What elements of performance appear in Standard Four?
19. Tell me about your understanding of and perceptions about Standard Four: Teachers facilitate learning for their students.
20. How have you been rated under Standard Four and what type of feedback have you received?
21. Give some examples of how you have used feedback in this area to inform your practices.
22. What elements of performance appear in Standard Five?
23. Tell me about your understanding of and perceptions about Standard Five: Teachers reflect upon their practice.
24. How have you been rated under Standard Five and what type of feedback have you received?

25. Give some examples of how you have used feedback in this area to inform your practices.
26. Are their certain factors that make Standards One through Five unfair? Please explain.

Knowledge of Value-Added Measures

27. Tell me about your understanding of value-added data.
28. Tell me about your general perceptions of VAD and your thoughts about it being included in the NCEES.
29. Please describe your own value-added data from last year and how you were rated for this Standard. What reflections, if any, were you able to make as a result of the VAD you received?
30. How did your value-added data from last year compare to previous years? What do you think about this information?

Impact on Teacher Practice

Questions 9, 13, 17, 21, 25, and 29

Improving the Teacher Evaluation Process

31. Do you think the NCEES is a fair measure of teacher effectiveness? Please explain.
32. How do you think the NCEES or the teacher evaluation process could be improved?
33. In what ways do you believe teachers' effectiveness should be measured and teachers' performance be evaluated?

34. What changes, if any, in teacher evaluation do you believe need to be made in order to improve student achievement?

APPENDIX D

PARTICIPANT DEMOGRAPHICS

Participant	School	Years of Exp.	Subjects Taught	Grades Taught	Educational Level	Work Background
1	Scholz Elem.	3.5	Math, Reading, Science	5	Bachelors	Advertising Manager
2	Scholz Elem.	15	Math, Reading, Science	5	Masters	Teaching in NC Only
3	Scholz Elem.	2.5	Math, Reading, Science	4	Doctorate	Lawyer
4	Scholz Elem.	14	Reading, Science, Social Studies	5	Bachelors	Business Owner, Educational Consultant
5	Sheehan Middle	16	Math	8	Masters	Human Resources
6	Delp Magnet	5	Reading, Social Studies	6	Masters	Teaching in NC Only
7	Delp Magnet	5	Reading	4	Masters	Teaching in NC Only
8	Sheehan Middle	8	Reading, Science	6, 7, 8	Bachelors	Teaching in NC Only
9	Masdea Middle	18	Science	6, 7, 8	Doctorate	Out of State Experience
10	Masdea Middle	22	Reading	7	Bachelors	Teaching in NC Only
11	Masdea Middle	8	Reading	7	Masters	Worked in a church
12	Sheehan Middle	9	Science, Social Studies	6	Masters	Police Officer

Participant	School	Years of Exp.	Subjects Taught	Grades Taught	Educational Level	Work Background
13	Masdea Middle	15	Math, Reading, Social Studies	6, 7, 8	Masters	Teaching in NC Only
14	Masdea Middle	15	Social Studies	7, 8	Bachelors	Business Owner
15	Sheehan Middle	9	Reading	6	Masters	Construction
16	Masdea Middle	2	Reading	8	Masters	Teaching in NC Only
17	Delp Magnet	11	Reading	7, 8	Bachelors	Teaching in NC Only
18	Goudreau Elem.	6	Reading, Math	5	Masters	Out of State Experience
19	Goudreau Elem.	19 or 20	Math, Science	4	Masters	Out of State Experience
20	Hashian Elem.	1	Math, Reading, Social Studies	5	Bachelors	Teaching in NC Only
21	Masdea Middle	3	Science	7	Masters	Teaching in NC Only
22	Goudreau Elem.	9	Reading, Math, Social Studies	5	Masters	Out of State Experience
23	Hashian Elem.	6	Reading, Math, Science, Social Studies	4	Bachelors	Out of State Experience

APPENDIX E

PARTICIPANT PERCEPTIONS OF STANDARD ONE

Standard One: Teachers Demonstrate Leadership																				
Participant	Teachers lead in their classroom.				Teachers demonstrate leadership in the school.				Teachers lead the teaching profession.				Teachers advocate for schools and students.				Teachers demonstrate high ethical standards.			
	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.
1																				
2						X	X					X								
3																				
4																				
5						X	X	X												
6							X													
7							X													
8		X				X	X													
9					X	X	X													
10							X	X												
11						X														
12						X														
13																				
14							X													
15												X								
16												X								
17							X													
18						X														
19							X													
20						X														
21						X														
22							X													
23						X														

APPENDIX F

PARTICIPANT PERCEPTIONS OF STANDARD TWO

Standard Two: Teachers Establish a Respectful Environment for a Diverse Population of Students																				
Participant	Teachers provide an environment in which each child has a positive, nurturing relationship with caring adults.				Teachers embrace diversity in the school community and in the world.				Teachers treat students as individuals.				Teachers adapt their teaching for the benefit of students with special needs.				Teachers work collaboratively with the families and significant adults in the lives of their students.			
	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.
1														X						
2		X				X														
3		X				X														
4			X						X									X		
5		X																		
6																				
7																				
8							X													
9							X													
10						X														
11		X																		
12																				
13		X				X														
14		X																		
15							X													
16						X														
17		X																		
18		X																X		
19						X														
20					X															
21		X																		
22			X			X														
23																				

APPENDIX G

PARTICIPANT PERCEPTIONS OF STANDARD THREE

Standard Three: Teachers Know the Content They Teach																
Participant	Teachers align their instruction with the North Carolina Standard Course of Study.				Teachers know the content appropriate to their teaching specialty.				Teachers recognize the interconnectedness of content areas/disciplines.				Teachers make instruction relevant to students.			
	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.
1																
2	X															
3														X		
4	X									X						
5	X									X	X					
6											X					
7																
8	X															
9										X						
10						X										
11																
12																
13																
14	X															
15																
16																
17														X		
18		X														
19																
20		X														
21																
22																
23																

APPENDIX H

PARTICIPANT PERCEPTIONS OF STANDARD FOUR

Standard Four: Teachers Facilitate Learning for Their Students																
Participant	Teachers know the ways in which learning takes place, and they know the appropriate levels of intellectual, physical, social, and emotional development of their students.				Teachers plan instruction appropriate for their students.				Teachers use a variety of instructional methods.				Teachers integrate and utilize technology in their instruction.			
	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.
1										X						
2		X														
3	X															
4	X															
5																
6																
7																
8																
9	X															
10																
11						X					X					
12						X				X						
13																
14																
15																
16																
17			X													
18																
19																
20		X														
21																
22										X						
23																

Participant Perceptions of Standard Four, Continued

Standard Four: Teachers Facilitate Learning for Their Students																
Participant	Teachers help students develop critical-thinking and problem-solving skills.				Teachers help students work in teams and develop leadership qualities.				Teachers communicate effectively.				Teachers use a variety of methods to assess what each student has learned.			
	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.
1																
2																
3																
4		X					X						X			
5							X							X		
6																
7																
8															X	
9																
10													X			
11																
12														X		
13																
14																
15																
16																
17																
18																
19																
20																
21																
22									X							
23														X		

APPENDIX I

PARTICIPANT PERCEPTIONS OF STANDARD FIVE

Standard Five: Teachers Reflect on Their Practice												
Participant	Teachers analyze student learning.				Teachers link professional growth to their professional goals.				Teachers function effectively in a complex, dynamic environment.			
	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.	Dev.	Prof.	Acc.	Dist.
1												
2												
3												
4												
5						X						
6						X						
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												

APPENDIX J

PARTICIPANTS' VIEWS ON FAIRNESS OF STANDARDS-BASED EVALUATION

Participant	Fair	Unfair					
		Subjective	Items Not Observable	Puts Some Teachers at Disadvantage	VAD Influences Observation Ratings	Only a Snapshot	Negative Relationship with Evaluator
1				X			
2						X	
3		X	X	X			
4		X					
5		X					
6				X			
7		X					
8		X		X			
9			X				
10				X			
11				X			
12						X	X
13							
14	X						
15					X	X	
16	X						
17	X						
18			X				
19							X
20		X	X				
21	X						
22		X	X				
23			X				
Totals	4	7	6	6	1	3	2

APPENDIX K

PARTICIPANT RESPONSES ON THE USEFULNESS OF VAD FOR REFLECTIVE PURPOSES

Responses on Usefulness of VAD			
Participant	**Useful for Reflection**	**Cannot Recall Enough to Reflect**	**VAD Not Useful for Reflection**
1	X		
2		X	
3		X	
4		X	
5		X	
6	X		
7			X
8			X
9		X	
10			X
11			X
12		X	
13		X	
14		X	
15			X
16			X
17			X
18	X		
19		X	
20		X	
21			X
22		X	
23		X	
Totals	3	12	8