

## ABSTRACT

LAMPRON, PHILLIP ANDREW. A Phenomenological Study of Consistency between Objective and Subjective Measures of Teacher Performance (Under the direction of Dr. Lance D. Fusarelli and Dr. Michael E. Ward).

Accountability in public education can often be a polarizing topic of discussion. Public schools are funded by public money, and thus have a responsibility to the public to ensure that children are receiving a proper education (Danielson, 2012). Teacher evaluations are one way that such accountability is implemented. In North Carolina teachers are evaluated using both a subjective measure called the North Carolina Educational Evaluation System (NCEES) and an objective measure called the Education Value-Added Assessment System (EVAAS). In this study I examined the perceptions of middle grades principals about the correlations between those measures.

The participants in this study were located in the Northeast and North Central regions of North Carolina and six of these seven principals worked at schools where greater than 50% of the students received free or reduced priced lunches. Participants were provided with graphs from the North Carolina Department of Public Instruction (NCDPI) that displayed their districts' correlations between ratings for middle grades mathematics teachers on Standards 1 and 4 in NCEES and their districts' ratings on EVAAS for those same teachers. For both Standard 1 and 4, weak correlations existed between the NCEES ratings and EVAAS ratings of those same teachers. Correlations were slightly stronger towards the low and high extremes of both scales.

In the review of archival data, this study found weak correlations between NCEES and EVAAS ratings of middle grades mathematics teachers in participating districts. These results align with findings from similar studies conducted in recent years (Batten et al., 2012; Henry & Guthrie, 2015). Each participant reviewed these data at the beginning of a conversation in which

I used an original interview protocol to solicit their feedback on these correlations and their perceptions about such correlations. This study found that participants had received very little training on either evaluation method in recent years and that a consistent understanding of how to implement these measures did not exist among the participants. Some of the leaders in the districts where these principals served had more conversations with participating principals about how alignment should exist between the two evaluation measures, but there was a lack of concrete examples regarding how to make sure that such alignment happened.

Based on these results, it is recommended that policymakers and leaders in education should reevaluate the two measures by which teachers are evaluated in North Carolina to see if adjustments can be made in order to create more alignment between these two systems. Policymakers could also evaluate the use of EVAAS ratings as they pertain to teacher ratings. The students of North Carolina deserve well-informed teachers who can provide them with strong potential for academic gains (Hostetler et al., 2007).

The study concludes with potential implications for policy and practice based on the results of this study. Suggestions for future research are also described in Chapter 5.

© Copyright 2024 by Phillip Andrew Lampron

A Phenomenological Study of Consistency between Objective and Subjective Measures of  
Teacher Performance

by  
Phillip Andrew Lampron

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Education

Educational Leadership

Raleigh, North Carolina  
2024

APPROVED BY:

---

Dr. Lance D. Fusarelli  
Committee Co-Chair

---

Dr. Michael E. Ward  
Committee Co-Chair

---

Dr. Anna J. Egalite

---

Dr. Timothy A. Drake

## DEDICATION

To Luca,

You are an inspiration to me in more ways than I know how to explain. Your joy for life reminds me to smile through difficult times. Your laugh has the ability to bring me joy no matter what else is happening. You're my favorite guy in the world, buddy.

I don't actually know if you'll ever read this, but in case you do, I hope that you choose to pursue whatever passion God lays on your heart. Once you find what that passion is, keep moving towards it day after day. You may not make huge strides each day, but as long as you keep moving forward, even if it's just one step, you're still going in the right direction. The only way you ever truly fail in life is if you quit. If you keep moving, you haven't failed. You might fall a few times, but life isn't about how many times you fall. It's about how many times you get back up (Proverbs 24:16).

I wish I knew how to tell you just what you mean to me, but I hope that I can show you throughout your life. I love you, son. I love you, I love you, I love you.

Love,

Dad

## **BIOGRAPHY**

After completing his bachelor's degree in elementary education at the University of North Carolina at Wilmington, Phillip began his career teaching in Pender County, North Carolina, in December 2007. In July 2009, he transitioned to Edgecombe County where he taught for two years before becoming a data coach for the district. He then transitioned into an instructional technology facilitator role, followed by becoming an instructional coach. He then went into the role of assistant principal at an elementary school. After that, Phillip served as the Ready Schools Specialist at the Down East Partnership for Children before becoming the K-7 Dean of Academics at North East Carolina Preparatory School (NECP) in Tarboro, North Carolina. He then became the upper school principal for NECP and served in that capacity for four years before stepping down to spend time with his son in June 2021. After approximately 18 months of entrepreneurship and focusing on being a dad, Phillip returned to education as a teacher, and he currently teaches 8<sup>th</sup> grade mathematics at South Edgecombe Middle School. Phillip also holds a master's degree in curriculum and instruction from Arkansas State University where he also completed the requirements for a principal add-on license.

## ACKNOWLEDGMENTS

I'd like to first acknowledge my Lord, my God, my savior, Jesus Christ. Throughout my life I have gone through some struggles with deciding what I wanted my relationship with Him to look like. I feel like I have finally come to grips with things that have happened in my life and my foundation is firmly planted on the rock, my cornerstone.

I'd also like to give a special acknowledgement to my mom. She's proven time and time again that she will be there for me in my most difficult times. She's gone through a lot with me, and her love and compassion have never wavered. She also encouraged me to complete this study with such poetic words as "I'd like to see you finish that before I die." Thanks, Ma. I love you.

I'd like to give a sincere acknowledgement to Lauren, as well. We've gone through quite a bit, but regardless, you encouraged me to start this process and I appreciate that. You're also a wonderful mother to our son and that means a great deal to me.

To my committee: Dr. Lance Fusarelli (co-chair), Dr. Anna Egalite, and Dr. Timothy Drake, I truly appreciate your diligence and willingness to stay with me through this process. I know it was a lot longer than expected, but you're all still here! It means a lot to me that I still get to have your signatures on my dissertation. It means more than you know.

To Dr. Ward, my co-chair, I really don't think I have the words to express my appreciation for you. You've been part of my life for approximately 9 ½ years now, and I hope that I get to continue counting up. You are such an incredible person, and you have been to each of us whom you have supported through our educational endeavors. I've never met someone who had anything but positive remarks for you. Of course, you've been a consistent source of feedback and guidance throughout the process of creating this dissertation, but you've also been

a consistent source of stability for me. You embody the characteristics that I hope others will see in me one day, the characteristics that I hope to pass on to Luca. You held Luca the day Lauren walked across the stage, and we'll recreate that picture here in just a few weeks. I lost my grandfather during the middle of this process, and I'm not saying that I see you as my grandfather, you're not that old (lol), but I do see in you some of the things I miss about him. I know in my heart of hearts that one of the main catalysts for me completing this study now is because you told me you're retiring at the end of this semester. I couldn't bear the thought of having someone else's name on this work.

There are several other friends and family members who have provided encouragement along the way, likely too many to list. There are a few individuals, however, who stepped up for me in the past year or so in ways that show me what true friendship means. My mom and stepdad, my brother Danny, little "brothers of choice" - Isaac and Aaron, Billy and Denesha, Julian and Bev, Patrick and Amy, Matt and Jessica, you've all drawn closer to me in my time of need and for that I will be eternally indebted. Additionally, to the rest of the "Sunday Crew," Jeff and Jennifer, Stephen and Inez, Tim and Kelly-Anne, our Sunday afternoon gatherings have given me something to look forward to as I was closing out the previous week and beginning a new one. I knew each Sunday someone was likely going to ask me "how's that dissertation going?" Knowing that this question was going to come up helped to keep me continuing down this path to completion. To my dad and stepmom, I truly appreciate all of your love and support.

I love you all.



## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
Chapter Introduction .....	1
Statement of the Problem.....	3
Purpose of the Study .....	7
Research Questions.....	8
Definition of Terms.....	9
Significance of the Study .....	11
Organization of the Study .....	13
Chapter Summary .....	14
<b>CHAPTER 2: LITERATURE REVIEW .....</b>	<b>15</b>
Chapter Introduction .....	15
Background and Policy Context .....	15
Overview of Educator Evaluations .....	15
Subjective Educator Evaluations in North Carolina .....	20
EVAAS Reports in North Carolina.....	24
Theoretical Framework.....	26
Pertinent Research and Expert Perspectives .....	29
Objective Evaluation and Value-Added Modeling (VAMs) .....	29
Subjective Evaluation Measures .....	34
Debate over the Necessity of Both Subjective and Objective Evaluations.....	37
Chapter Summary .....	38
<b>CHAPTER 3: METHODOLOGY .....</b>	<b>39</b>
Chapter Introduction .....	39
Research Design.....	40
Research Questions and Hypotheses .....	41
Study Participants .....	42
Participant Descriptions .....	42
Variables .....	44
Instrumentation .....	48
Procedures.....	50
Data Collection .....	51
Data Analysis .....	52
Subjectivity Statement .....	54
Delimitations of the Study .....	55
Assumptions.....	55
Chapter Summary .....	56
<b>CHAPTER 4: RESULTS .....</b>	<b>58</b>
Chapter Introduction .....	58

Participant Profiles.....	60
Research Questions.....	63
Chapter Summary .....	98
<b>CHAPTER 5: DISCUSSION .....</b>	<b>99</b>
Chapter Introduction .....	99
Purpose of the Study .....	99
Summary of Findings.....	100
Discussion of Findings.....	104
Discussion of Quantitative Findings.....	105
Discussion of Qualitative Findings.....	107
Reflections on Theoretical Framework.....	113
Limitations of the Study.....	115
Implications of the Study for Policy and Practice .....	116
Implications for Policy.....	117
Implications for Practice .....	118
Recommendations for Future Research .....	120
Chapter Summary .....	121
<b>REFERENCES .....</b>	<b>124</b>
<b>APPENDICES .....</b>	<b>142</b>
Appendix A: Protocol for Principal Interview.....	143
Appendix B: Principal Perceptions of NCEES and Principal Predictions of Teachers' EVAAS Ratings Validity Questionnaire.....	146
Appendix C: Superintendent Consent Form.....	149
Appendix D: Principal Consent Form.....	152

## LIST OF TABLES

Table 4.1 – Participant Demographics.....	63
---	----

## LIST OF FIGURES

Figure 4.1 .....	66
Figure 4.2 .....	68
Figure 4.3 .....	70
Figure 4.4 .....	73
Figure 4.5 .....	76
Figure 4.6 .....	78
Figure 4.7 .....	81
Figure 4.8 .....	83

## **CHAPTER 1: INTRODUCTION**

### **Chapter Introduction**

The purpose of this study was to examine the perceptions of principals in the Northeast and North Central regions of North Carolina regarding the relationships among the Educator Value-Added Assessment System (EVAAS) ratings and the North Carolina Educator Evaluation System (NCEES) ratings of middle grades mathematics teachers. Value-added assessment models (VAMs) have been discussed frequently in recent years due to their application in the evaluations of both schools and teachers. In North Carolina, each teacher, administrator, and school receives a rating of does not meet expected growth, meets expected growth, or exceeds expected growth based on calculations within EVAAS, which was created by the SAS Institute (SAS Institute, Inc., 2016a). During the 2013-2014 and 2014-2015 school years, teachers received one of these designations under a separate standard as a part of their ratings in the North Carolina Educator Evaluation System (NCEES) (Public Schools of North Carolina, 2013). EVAAS ratings can now be used in conjunction with the other five standards in NCEES to make decisions about educators, but EVAAS ratings are no longer issued as a stand-alone measure of effectiveness (North Carolina State Board of Education, 2016).

While it is not explicitly stated in the literature, it is reasonable to infer that if decisions about how to use teacher effectiveness scores from EVAAS are left to the discretion of the principal and local education agency (LEA), then the principal and LEA may elect not to use these results at all when making personnel decisions. EVAAS ratings are still used as a part of each school's School Performance Grade (SPG) (North Carolina Department of Public Instruction, 2017b). SPGs for each school consist of a combination of the school's overall proficiency and their growth score per EVAAS. The SPG is calculated using a ratio of 80% of

the score coming from the proficiency score of the school, and 20% from their EVAAS growth score (North Carolina Department of Public Instruction, 2015a). However, at the time of this study, the North Carolina House of Representatives had passed a bill that would change that ratio to 50/50. The bill was being reviewed by the Senate (North Carolina Association of School Administrators, 2017). This proposed legislation (House Bill 322) also stated that if a school's School Performance Grade is lowered by adding in the growth metric, then the school can elect to only use the proficiency scores to calculate the grade. The legislation did not pass in the Senate and the 80/20 ratio for proficiency and growth remains in effect.

The policy changes that shifted the use of EVAAS scores from mandatory to optional in teacher evaluation processes were enacted in April 2016 by the State Board of Education. A teacher's rating in EVAAS is now used as an additional element of information when rating teachers on Standards 1-5 of the NCEES (North Carolina State Board of Education, 2016). Performance relative to Standard 6, which is documented through the EVAAS metric, would be considered an objective measure of teacher effectiveness based on the definition Jacob and Lefgren (2008) used when describing other VAMs. Standards 1-5 of the NCEES for teachers are subjective in nature, according to the definition from the same research by Jacob and Lefgren, and evaluative ratings of performance relative to these standards are based on a rubric that contains designations that range from not demonstrated to distinguished in the following areas: Standard 1, teachers demonstrate leadership; Standard 2, teachers establish a respectful environment for a diverse population of students; Standard 3, teachers know the content they teach; Standard 4, teachers facilitate learning for their students; Standard 5, teachers reflect on their practice (McREL, 2015). Teachers receive ratings in these areas based on where an evaluator, usually an administrator, rates them along the continuum of not demonstrated to

distinguished, using criteria provided by the state (McREL, 2015). Since the administrator exercises judgment to determine whether or not a teacher has sufficiently met the rating criteria, these evaluation judgments can be deemed subjective.

The remaining sections of this chapter serve to introduce the study of the relationship between the EVAAS scores of teachers and their ratings from administrators on the NCEES. The Statement of the Problem addresses the concerns regarding discrepancies between the ratings of teachers when using both subjective and objective measures. The Purpose of the Study section gives details about how the study addressed these discrepancies. The Research Questions section provides insight into what I specifically studied. The Definition of Terms section provides insight to the reader with regards to how certain terms are defined in this study. The Significance of the Study section offers reasoning for the need for such a study. The Organization of the Study section gives a description of how the study was formatted, and the Summary gives an overview of what I presented in Chapter 1.

### **Statement of the Problem**

It is important that educators receive accurate feedback from evaluations, whether evaluation protocols are subjective or objective in nature. Public schools are funded with public money, and thus it is the right of the taxpayers to expect that educators in these classrooms be effective practitioners (Danielson, 2012). The evaluations that teachers receive should be based on multiple, valid indicators so that stakeholders can be assured that they are reliable and beneficial (Danielson, 2012). Contemporary educator evaluations are often viewed as inconsistent and unreliable (Darling-Hammond, 2013), and rarely differentiate between the teachers with varying degrees of effectiveness (Sartain et al., 2011). Current teacher evaluation processes have also increased considerably the amount of time principals must spend with each

individual educator in order to complete the evaluation process (Kersten & Israel, 2005; Weisberg et al., 2009). Recent evaluation results in the Northeast and North Central regions of North Carolina appear to indicate that a “widget effect,” a term described by Weisberg et al. (2009), is present. The “widget effect” refers to “the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher” (p. 4). Educator evaluation results in the region indicate that the vast majority of teachers are rated at the top three performance levels of the state’s evaluation Standards 1-5; these three performance levels are proficient, accomplished, and distinguished. Ratings based on their EVAAS scores under Standard 6, however, are much more variable (North Carolina School Report Cards, 2015).

The conversation about using VAMs in education started getting attention in the early 1980s when William Sanders began creating what is now known as the Tennessee Value-Added Assessment System (Stewart, 2006). Prior to that, the use of such measures had been relegated to the private sector (Lewis & Fusarelli, 2010). North Carolina began officially using the Education Value-Added Assessment System (EVAAS) as a measure of teacher effectiveness in 2012 (Rose et al., 2012). One of the reasons North Carolina moved forward with using VAMs in teacher evaluations was to meet the requirements of the Race to the Top (RttT) Grant (U.S. Department of Education, 2009). North Carolina applied for the RttT Grant that was approved and award in 2010 (Marks et al., 2015). The RttT Grant required that student growth measures be included in teacher evaluations (U.S. Department of Education, 2009). In total, 45 states applied for the RttT Grant, and 19 states were approved (Miller & Hanna, 2014). All of the approved states were required to have some form of student growth factored into teacher evaluations (U.S. Department of Education, 2009).



The adoption of such measures on a national scale did not come without scrutiny. VAMs are considered unstable by some researchers (Baker et al., 2010; Everson, 2017), and others also assert the presence of sorting bias, since students are not randomly assigned to teachers (Paufler & Amrein-Beardsley, 2014; Sass et al., 2014). Still others are concerned that VAMs show little to no correlation with subjective measures of teacher evaluation (Amrein-Beardsley et al., 2016).

As has been the case in many other states, North Carolina has adopted both subjective (McREL, 2015) and objective (North Carolina Department of Public Instruction, 2016a) measures of teacher evaluation. The NCEES is subjective in nature according to the definition provided by Jacob and Lefgren (2008), and teacher effectiveness scores in EVAAS are considered objective according to the same study. Subjective evaluation ratings can be biased depending on the leniency of the evaluator within each standard (Rockoff & Speroni, 2010). Some researchers contend that VAMs such as EVAAS do not take into account a variety of factors that influence student achievement. They are adamant that VAM scores should not be considered when making decisions regarding a teacher's employment or their pay levels (AERA Council, 2015). Proponents of VAMs, on the other hand, assert that VAMs have the ability to improve current accountability systems and provide beneficial feedback to administrators (Harris, 2009; Scherrer, 2012). Questions arise as to which type of evaluation is more precise, which more accurately indicates a teacher's impact on his/her students, and which is more effective in prompting positive changes in teacher practice.

Within EVAAS, there are a variety of reports that teachers and administrators can view to gauge the success of students and teachers within their schools. The report that becomes part of a teacher's evaluation is entitled Teacher Effectiveness Report (North Carolina Department of Public Instruction, 2016b). This is a relatively complex calculation that seeks to determine if

students within a teacher's classroom reached at least one year's worth of academic growth during a school year. Based on the results of these calculations, each teacher is rated at one of the following EVAAS performance levels: does not meet expected growth, meets expected growth, or exceeds expected growth. This type of reporting has been in place in North Carolina for multiple years, yet many teachers and administrators are relatively unfamiliar with what EVAAS reporting actually shows them (Hewitt, 2015). They are likewise ill-informed about the impact that it could have upon their school, and on them personally (Hewitt, 2015). If policymakers, administrators, and teachers are going to use these results to make decisions regarding student placements or personnel decisions, then they should have a firm grasp of what reports within EVAAS mean. Adequate comprehension of such reports is difficult, however, considering the complex mathematical equations that are used to create these reports (Amrein-Beardsley, 2008).

There are many factors that play a part in a child's education. It is therefore reasonable to consider that multiple factors should be taken into consideration when evaluating the impact of a teacher on her/his students. While there are a variety of VAMs that can provide accurate and beneficial statistical information to schools, many researchers contend that these data should be used in conjunction with data from classroom observations when teachers are evaluated (Baker et al., 2010; Scherrer, 2012; Hill et al., 2010; Xu et al., 2016).

Current educator evaluations often fall short in actually evaluating the quality of education that a teacher provides to her/his students (Hill et al., 2010). In some instances, they have also been found to have little impact on student achievement (Bleiberg et al. 2024). There are some subjective teacher evaluation protocols that exist, however, that seem to provide a stronger correlation with increased student achievement. Such subjective observations as the Protocol for Language Arts Teaching Observation (PLATO) (Grossman et al., 2014),

mathematical knowledge for teaching (MKT), mathematical quality of instruction (MQI) (Hill et al., 2010), and the Rapid Assessment of Teacher Effectiveness (RATE) (Gargani & Strong, 2014) may more accurately align with value-added measures of teacher effectiveness. It is imperative that sound research expand what is known about the degree to which VAMs are adequate and accurate predictors of teacher effectiveness, that disconnects between teacher impact as measured by VAMS and as measured by subjective protocols be more sufficiently examined, and that models of teacher evaluation be adapted to comport with the findings of such research. To do otherwise is to allow concerns such as those described in 2010 by Baker et al. that the scores on VAMs are unstable and lack validity to persist and, worse, risk sending inaccurate signals to teachers about the nature of their performance.

### **Purpose of the Study**

This study used a qualitative design to gain insights from middle grades principals regarding how middle grades mathematics teachers' subjective evaluations in the NCEES are related to their objective ratings in the Teacher Effectiveness Report in EVAAS. This study further sought to determine if the magnitude of and correlations between these constructs are related. NCEES ratings of middle grades mathematics teachers throughout the Northeast and North Central regions of North Carolina were compared to the EVAAS ratings of these same individuals. Principals in this region were interviewed to gain insight into their perspectives regarding correlations between NCEES and EVAAS ratings of teachers.

A substantial body of literature discusses either methodological concerns or support for VAMs. The same is true of subjective teacher evaluations being studied in isolation (Amrein-Beardsley, 2008; Collins, 2014; Danielson, 2012; Dowley & Kaplan, 2014; Franco & Seidel, 2012). The literature is not as expansive when it comes to discussing correlations between these

two measures of teacher effectiveness, particularly with regards to EVAAS and NCEES correlations. This study sought to determine if correlations between these two types of evaluations exist, and to what degree. This information can inform future teacher evaluation methods, and aid in the determination of where professional development should be focused to impact the professional development of administrators in order to improve their effectiveness as evaluators of teacher practice.

In Chapter 3 of this dissertation, the qualitative methods that were employed in this study are described. These methods consisted of an interview protocol that was conducted with principals in the Northeast and North Central regions of North Carolina. Principals were asked to provide insight regarding their training on both evaluation measures, and their opinions on how these measures correlate with each other and with what they see in their school buildings on a day-to-day basis.

### **Research Questions**

The research questions for this study are as follows:

1. How do the correlations of NCEES and EVAAS ratings compare among the participating districts in the Northeast and North Central regions of North Carolina?
2. How do principals describe the levels of correlation that exist between the NCEES and EVAAS ratings of middle grades mathematics teachers in their district?
3. How do principals describe the levels of correlation that exist between their teachers' NCEES ratings and EVAAS ratings and those of the district?
4. To what factors do principals in the Northeast and North Central regions of North Carolina attribute the level of correlation between the NCEES and EVAAS ratings of teachers?

5. What teacher attributes do principals in the Northeast and North Central Regions of North Carolina believe would increase student achievement and thus produce a stronger correlation in NCEES and EVAAS ratings, if an evaluation of these attributes existed in the NCEES?

### **Definition of Terms**

The terms listed below are used frequently in this study and are defined as follows:

*Cognitive evaluation theory* – This theory asserts that if employees are intrinsically motivated by their occupation, they are more likely to be successful in that occupation (Deci et al., 1999).

*EVAAS* – This stands for the Education Value Added Assessment System. EVAAS is used to determine the growth factor for schools in North Carolina (North Carolina Department of Public Instruction, 2016c). EVAAS is a value-added assessment model that helps schools make data-based decisions about their institution (SAS Institute, Inc., 2016a).

*Expertise theory* – In this theory, the researchers found that those who were experts in their field progressed along a continuum from novice to expert as they continued to practice in their respective roles. The skills that they possessed and needed to progress to the next level varied along the continuum (Ericsson et al., 1993).

*Highly qualified teacher* – This term originated with the No Child Left Behind legislation (No Child Left Behind Act of 2001). During that time, this term meant that a teacher had at least two years of experience in a classroom setting, and a postsecondary education in the field in which they taught. Since December 2015, states have had the flexibility to determine what they would like for this term to mean with the institution of the Every Student Succeeds Act (U.S. Department of Education, 2015).

*NCEES* – The North Carolina Educator Evaluation System. This term is used to describe the evaluation process that teachers in North Carolina go through each year. During this process teachers are rated on the following five standards: Standard 1, teachers demonstrate leadership; Standard 2, teachers establish a respectful environment for a diverse population of students; Standard 3, teachers know the content they teach; Standard 4, teachers facilitate learning for their students; Standard 5, teachers reflect on their practice. Within these five standards, teachers are rated along a continuum from not demonstrated to distinguished based on the observations administrators and/or peer observers perform throughout the school year (North Carolina Department of Public Instruction, 2016a). This type of evaluation would be considered subjective.

*Normal curve equivalents (NCEs)* – This term is used to describe the scale used to determine student growth for grades 3-8 reading and math assessments in North Carolina. Student scale scores are converted to NCEs because NCEs are on an equal-interval scale so student scores can be compared from year to year regardless of the assessment that they are taking (SAS Institute, Inc., 2016b).

*Objective evaluations* – These evaluations in education refer to a measure of a teacher's effectiveness based solely on outcomes of assessment data, typically standardized test data that has been analyzed using value-added modeling (Jacob & Lefgren, 2008).

*Social network theory* – This theory is based around the concept that when individuals believe that they are working together as a collaborative, cohesive unit, they are more likely to interact with one another and learn from each other's experiences (Liou et al., 2015). This speaks to the idea that when principals are placed in the role of evaluator, teachers may be less likely to receive feedback in a constructive manner.

*Subjective evaluations* – This term refers to evaluations of teachers by administrators where teachers are rated based on the administrator’s perception of their teaching style or ability. Often these ratings are based upon descriptions listed in a rubric (Jacob & Lefgren, 2008).

*Widget effect* – The “widget effect” refers to “the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher” (Weisberg et al., 2009, p. 4). To paraphrase, this occurs when teacher ratings look very similar regardless of individual teachers’ ability to impact student achievement.

### **Significance of the Study**

There were multiple justifications for a study of this sort. Teachers need to be provided with feedback about how to improve their craft so that they are better able to positively impact student achievement. School level administrators are in a unique position in the field of education. They are expected to build relationships with their staff members, ensure that each classroom has a qualified teacher, and yet still hold these staff members accountable. If educators are unhappy with the ratings that they receive from their supervisors, they may be encouraged to seek employment at another institution. This would leave the school at which they are currently employed with a vacant position. If schools are not fully staffed with highly qualified teachers according to the definition given by the NCDPI, the principals of those schools may be deemed as somewhat inferior in their hiring practices. The percentage of classrooms with highly qualified teachers is reported for each school on the School Report Cards from the NCDPI. If this percentage is low, the perception given to the public is that this school is not adequately prepared to educate the students they serve. These circumstances suggest that administrators have an abundance of reasons to inflate ratings in the NCEES in order to ensure that they are able to retain teachers that are deemed highly qualified by the NCDPI (Danielson, 2016).

The designation of highly qualified was originally given to teachers who had at least two years of experience in a classroom setting and a postsecondary education in the field in which they taught according to the No Child Left Behind Act (NCLB) (2001). When NCLB was replaced by the Every Student Succeeds Act (ESSA) in December 2015 (U.S. Department of Education, 2015), the terminology used in legislation was altered to allow states to use their own discretion regarding what they deemed to be a highly qualified teacher (Sawchuk, 2016). According to the North Carolina Department of Public Instruction (2015b), the requirements to be highly qualified vary depending on the grade span a teacher is teaching. The guidelines are listed as:

To be a Highly Qualified teacher at the elementary school level, a teacher must have obtained an appropriate license for the core academic subjects taught and demonstrate subject knowledge and teaching skills in reading/language arts, writing, mathematics, and other areas of the basic elementary school curriculum by passing the teacher licensing exams (Praxis II) required by the state.

To be a Highly Qualified teacher at the middle and high school levels, a teacher must have obtained a middle school or secondary license in a teaching area required for each teaching assignment and demonstrate a high level of competency by:

- Passing the required Praxis II test(s) in each academic subject in which the teacher teaches, **or**
- Successfully completing **one** of the following steps in each academic subject in which the teacher teaches:



1. An undergraduate major,
2. Coursework equivalent to an undergraduate major,
3. A graduate degree in the core teaching subject area(s),
4. Master's level licensure or above in the appropriate subject area, or
5. National Board Certification in the related subject area(s). (North Carolina Department of Public Instruction, 2015b)

This designation does not, in any way, reflect a teacher's performance on the NCEES, nor their Teacher Effectiveness rating in the EVAAS. While principals may have reason to attempt to retain teachers who meet this designation (Danielson, 2016), there is little evidence that shows that factors such as the ones outlined here have any significant impact on student achievement (Phillips, 2010).

### **Organization of the Study**

The balance of this study addresses the pertinent body of knowledge and the study procedures. The findings, conclusions, and related recommendations are addressed as well. Chapter 2 of this study contains a review of relevant literature pertaining to both subjective and objective evaluations. Historical context is provided to show how educator evaluations have progressed across time to now contain both principal ratings of educator effectiveness and value-added modeling. The theoretical framework that undergirds the study is also described in the second chapter. Chapter 3 presents information regarding the methodological approach of the study. A detailed description is given about sampling selection for this study, the survey distribution process, the interview protocol, and the data collection and disaggregation.

In Chapter 4, the results of the study are reviewed in detail. This chapter includes findings regarding the degree to which correlations were found to exist between NCEES ratings

of middle grades mathematics teachers and EVAAS ratings, as well as the perceptions of principals on the ratings between these two measures. Chapter 5 contains a discussion of the findings with suggestions for the policy, practice, and future research based on the results of the study.

### **Chapter Summary**

This chapter discussed the importance of educator evaluation systems that are both subjective and objective in nature. VAMs like the EVAAS and subjective ratings like the NCEES have been shown by some researchers to be poorly related (Amrein-Beardsley et al., 2016). The chapter also explored the idea that determining relationships between the NCEES and EVAAS could provide benefits to the field of education. Literature surrounding this topic was presented and other potential teacher evaluation models were mentioned; these included PLATO, MKT, MQI, and RATE. Information was also presented to indicate the importance of the study due to the need for improved accuracy in the provision of evaluation ratings to teachers which, in turn, have implications for their practice and effectiveness.

## **CHAPTER 2: LITERATURE REVIEW**

### **Chapter Introduction**

An expansive body of literature can be found on the topics addressed in this study. The concept of providing feedback to teachers regarding their performance dates back to the mid-1800s (Tracy, 1995). Educator evaluations today look much different than they did then. Teachers today receive not only anecdotal feedback, but statistical evaluations of their impact on student achievement. The following sections address this evolution in evaluation practice and explore the current policy environment surrounding such practices. A theoretical framework is described, and an examination of pertinent research and professional perspectives are provided.

### **Background and Policy Context**

In order to provide the reader with insights into the issues associated with teacher evaluation practice, the content that follows provides relevant historical developments in this area of practice. It is also useful to understand the policy environment in which the study occurred. Toward that end, relevant current policies that impact teacher evaluation practices are reviewed.

### **Overview of Educator Evaluations**

This study addresses two domains associated with educator performance assessment: subjective evaluations of teachers provided by administrators and objective evaluation ratings produced by value-added measures (VAMs) from statistical databases. While both processes can be controversial, debate about the latter has been particularly contentious. In North Carolina, the subjective evaluations are completed by supervisors based upon a rubric called the North Carolina Educator Evaluation System (NCEES). Data for the objective measures are produced by the Education Value-Added Assessment System (EVAAS). This dissertation seeks to

determine if there are relationships among the results yielded by these components of the evaluation system, as well as the perceptions of principals on these measures and what teacher attributes they feel have the potential to increase student achievement.

Public education, as a profession, has progressively placed more and more importance on teacher evaluations. In the 1700s, teaching was seen almost as a public service because of the benefit that it provided the community. The position of teacher was viewed much like that of individuals who worked for churches of the time (Tracy, 1995). The local governments or the clergy would select teachers from the community and provide feedback to those selected and placed in service. The feedback given to teachers of the time was quite varied (Tracy, 1995). It was not until the mid-1800s that pedagogy began to be recognized as a significant factor in the teaching process. Edward Thorndike and Ellwood Cubberley began incorporating data analysis into education to be able to scientifically measure the outcomes of schooling in the early 1900s (Cubberley, 1929). William Wetzel (1929), a contemporary, suggested using students' scores as a measure of teacher effectiveness. In the period immediately following World War II, school administrators were increasingly viewed as supervisors, and supervisory functions included observing teachers and providing feedback (Lewis & Leps, 1946).

Clinical supervision was introduced in the late 1950s; in this model, prospective teachers would spend time observing in a veteran teacher's classroom to improve their practice. Robert Goldhammer (1969) outlined what educators today know as a formal observation process, complete with pre- and post-conferences, as well as data analysis. In the early 1980s, Madeline Hunter created the seven-step lesson plan (Hunter, 1984) that began to be viewed as a model for how a classroom should operate (Burns, 2006). During this period, the RAND group studied the types of supervision and evaluations that were taking place in America's schools. The RAND

researchers found that the feedback being provided was not specific enough to assist teachers in improving their practice and that teachers actually desired more standardized approaches.

Teachers did not believe that principals were able to accurately evaluate them without a specific protocol in place (Wise et al., 1984). Charlotte Danielson provided a model for evaluation and feedback in 1996, and in her work, *Enhancing Professional Practice: A Framework for Teaching*, Danielson presents a rubric consisting of four categories for ratings (similar to that of the NCEES) (Danielson, 1996).

Early in the 21st century, the role of observations began to shift from supervision to a heavier emphasis on evaluation (Marzano et al., 2011). The timing of this shift coincided with the implementation of the No Child Left Behind Act of 2001 (NCLB). This legislation brought with it an increased focus on teacher quality and improving teacher performance (No Child Left Behind Act of 2001). The bill was signed into law in January 2002 with bipartisan support (Editorial Projects in Education Research Center, 2015). Soon after, researchers such as Tucker and Stronge (2005) began to assert that student achievement should be considered in teacher evaluations. Tucker and Stronge suggested that student gains be considered in the evaluation process in their book *Linking Teacher Evaluation and Student Learning*, which was published in 2005 (Tucker & Stronge, 2005).

The NCEES in its current form has been used for teacher evaluations since the 2011-2012 school year (North Carolina Department of Public Instruction, 2016a). However, according to Williams et al. (2010), this evaluation format has been used to some extent since 2007. Within the NCEES, teachers are evaluated on performance against five standards and multiple elements within each standard. Teachers are rated on each of these elements using one of the following five hierarchical classifications: not demonstrated, developing, proficient, accomplished, or

distinguished (McREL, 2015). The reasoning behind using this multi-dimensional, subjective measure to evaluate teachers is to better ensure that the evaluator is using a variety of sources to determine a teacher's rating. Teaching is a complex occupation and should be evaluated using more than one data source (Grossman et al., 2014). Among things that the NCEES, as well as other subjective evaluation systems, does not do is attempt to predict student outcomes as a result of the ratings on the elements through which teachers are judged. The theory of action associated with systems like the NCEES is that educators focus their work on these standards with the assumption that performance that earns higher ratings on these standards will result in improved student achievement (Gargani & Strong, 2014).

EVAAS results officially became part of the NCEES in 2011-2012 (North Carolina Department of Public Instruction, 2016b), but North Carolina had been providing EVAAS reports to schools for many years prior to the 2011-2012 school year. Beginning in the 2012-2013 school year, EVAAS was used as part of the accountability model for North Carolina schools to provide a statistical indicator of student achievement growth (North Carolina Department of Public Instruction, 2016b). Even though EVAAS has been a tool to provide insights into educational practice in North Carolina for multiple years, a recent study has shown that there are still many educators who are not comfortable with either their knowledge base or training regarding EVAAS reports (Hewitt, 2015). Some researchers believe that that is not surprising, given what they perceive is the lack of information provided by SAS as to how their reports are created (Amrein-Beardsley, 2008). EVAAS is a multi-variate statistical analysis model that takes a tremendous amount of data into account when determining a student's growth score (Wright et al., 2010). The metric of particular interest as judgments are made about teaching performance is the Teacher Effectiveness rating. Some of the data that are collected

when calculating a student's growth are not necessarily used when calculating the Teacher Effectiveness rating, such as student demographic information and socio-economic status. Many educators believe that these factors impact student achievement. The demographic makeup of the class and the impact that being surrounded by more qualified peers can have on student scores (Franco & Seidel, 2012) are still other factors that are not included in VAMs such as Teacher Effectiveness ratings in EVAAS. What EVAAS does, in its most basic sense, is compare a student's achievement on a standardized test against "all Prior Achievement that is measured by a quality assessment, such as the end-of-grade (EOG) or end-of-course (EOC) assessments" (SAS Institute, Inc., 2016a, "What is value-added assessment?", para. 2). This information is then compiled and a teacher is given one of three ratings on a hierarchical scale (does not meet expected growth, meets expected growth, or exceeds expected growth) based on his/her teacher effect size.

The NCEES and EVAAS create evaluative ratings for teachers through two very different methods. While many educators may currently have a negative view of EVAAS, value-added modeling can provide beneficial information to schools and districts (Amrein-Beardsley, 2008). This is also the case with subjective ratings based on observations performed by administrators and colleagues, assuming that they are provided by trained administrators who provide accurate and reliable assessments of performance (Rockoff & Speroni, 2010). Neither form of evaluation should be discounted. In order to provide adequate benefit to the field of education, these two rating systems should be meshed together to provide high-quality feedback to teachers and administrators (Harris & Sass, 2014).

## **Subjective Educator Evaluations in North Carolina**

Statewide performance evaluations of teachers in North Carolina date back to 1985 (North Carolina Department of Public Instruction, 2017a). This first statewide performance evaluation instrument was the Teacher Performance Appraisal Instrument (TPAI) and was developed by staff at the North Carolina Department of Public Instruction (NCDPI) in conjunction with researchers from the University of North Carolina at Chapel Hill (Holdzkom, 1991). When developing this instrument, the research team reviewed a variety of skills with four main criteria in mind prior to recommendations of skills to include. These four criteria were:

1. The skill should be observable.
2. The skill should be alterable.
3. The skill should have a demonstrated effect on student achievement, time on-task, or both.
4. The skill should have been validated by empirical research studies conducted in various grades and subjects. (Holdzkom, 1991, p. 783)

The team's research ultimately produced 28 skills on which to evaluate teachers. Once created, this instrument was evaluated by individuals from the University of Virginia, the University of Tennessee, and the University of Connecticut, and was found by the reviewers to be an appropriate measure to use when making decisions regarding teacher certification and tenure (Holdzkom, 1991). The TPAI was necessary to use when considering teacher certification since in 1985 teachers were not fully certified when they entered the classroom; instead, they entered under a "supervised apprenticeship" (Milner, 1991, p. 465). The TPAI received strong criticism from both Olson (1989) and Milner (1991) as they contended that the instrument focused too



much on procedural actions of teachers and not enough on a teacher's ability to expand student learning.

In accordance with the Excellent Schools Act of 1997, the TPAI was revised to include differentiated requirements for beginning and career status teachers. This revised version, known as the Teacher Performance Appraisal Instrument – Revised (TPAI-R) was initially implemented in June 2000, and was required to be in use statewide by June 2001 (North Carolina Department of Public Instruction, n.d.). Using the TPAI-R, principals rated teachers based on eight (8) major functions. Those functions were:

1. Management of Instructional Time
2. Management of Student Behavior
3. Instructional Presentation
4. Instructional Monitoring
5. Instructional Feedback
6. Facilitating Instruction
7. Communicating within the Educational Environment
8. Performing Non-Instructional Duties (TPAI-R Full Review – Experienced Teachers, n.d., p. 104)

Within each of these eight major functions, there were suggested specific elements for which the evaluator should look, but the teacher only received an overall rating for each of the eight major functions. In total, there were 43 suggested elements that could be reviewed for a teacher's ratings on the functions. They did not receive a rating on each individual element, but were rated as above standard, at standard, below standard, or unsatisfactory on each of the eight major functions. This was true for both beginning and career status teachers (Teacher

Performance Appraisal Instrument – Revised, n.d.; TPAI-R Full Review – Experienced Teachers, n.d.).

In the 2006-2007 school year, the NCDPI adopted what it referred to as 21<sup>st</sup> century standards for teacher evaluations (North Carolina Department of Public Instruction, 2017a), and in 2008 the North Carolina State Board of Education (NCSBE) approved the use of the Rubric for Evaluating North Carolina Teachers and the Teacher Evaluation Process (McREL, 2015). This is the same rubric that is now used in what educators refer to as the North Carolina Educator Evaluation System (NCEES) (McREL, 2015). With this shift, administrators went from rating teachers on eight major functions with 43 guiding elements in the TPAI-R to rating teachers on five overarching standards, with 25 elements under those five standards, and reviewed 147 specific skills to complete a single evaluation with the NCEES (McREL, 2015). The five overarching standards on which teachers are rated in this rubric are:

Standard 1: Teachers Demonstrate Leadership

Standard 2: Teachers Establish a Respectful Environment for a Diverse Population of Students

Standard 3: Teachers Know the Content They Teach

Standard 4: Teachers Facilitate Learning for Their Students

Standard 5: Teachers Reflect on Their Practice (McREL, 2015, p. 7-11)

In 2011, the NCSBE updated their policy to include a component for teachers to be placed on an abbreviated cycle, and in 2015 the terms referencing career status and probationary teachers were eliminated (McREL, 2015).

With these changes to the NCEES, not all teachers are rated on all five standards every year, and not all teachers will receive the same amount of observations for the ratings to be based

on (McREL, 2015). Teachers that have three (3) years of experience or less are placed on the comprehensive cycle. A teacher on the comprehensive cycle will receive three observations each year from the principal or the principal's designee, as well as one peer observation from a fellow teacher. One observation during the school year must be done by the principal. All observations conducted on a teacher on the comprehensive cycle must be "formal" observations. That means that each observation must last at least 45 minutes, or an entire class period, and that a post-conference be conducted with the teacher after the observation is complete. The first formal observation must also be completed after a pre-conference has been conducted. In the pre-conference, the principal or principal's designee will review the teacher's self-assessment with them and discuss the upcoming observation (McREL, 2015).

Teachers that have greater than three years of experience can be placed on the comprehensive cycle if the school administrators feel that it is necessary, but they may also be placed on either a standard or abbreviated cycle. A teacher on the standard cycle will receive one formal observation with a pre-conference, and two more observation that can be formal or informal. Informal observations must only last 20 minutes, and a post-conference is not required with informal observations, but a teacher may request that one be performed. A teacher on the abbreviated cycle must only be observed twice throughout the school year, and both of those observations can be informal. Teachers on the abbreviated cycle are also only rated on standards 1 and 4. All teachers, regardless of their cycle, also receive a Summative Evaluation at the end of the school year. The Summative Evaluation is a compilation of the ratings that the teacher has received throughout the school year. Again, for teachers on the abbreviated cycle, their summative evaluation will only be on standards 1 and 4 (McREL, 2015).

## **EVAAS Reports in North Carolina**

Value-added measures were not always used to the degree that they now are employed in public education in North Carolina. From the time that NCLB was signed into law in 2002 states struggled to meet the Adequate Yearly Progress requirements outlined in the legislation (Editorial Projects in Education Research Center, 2015). In 2011, many states were still struggling to meet these standards, so the federal government offered waivers for NCLB for those states. If states applied for a waiver, they had to accept certain regulations in lieu of the requirements present in NCLB. One of these requirements was that student progress on standardized assessments become part of the teacher evaluation system (Editorial Projects in Education Research Center, 2015). North Carolina was granted such a waiver in May 2012 (North Carolina ESEA Flexibility Request, 2012). Since that time, EVAAS results have been used in the North Carolina accountability model (North Carolina Department of Public Instruction, 2016b).

There are multiple reports in the Educational Value-Added Assessment System (EVAAS) from which educators in North Carolina can gain insight for decisions being made in their schools. Some examples include School and District Academic Preparedness Reports, School and District Value-Added Reports, and School and District Diagnostic Reports (SAS Institute, Inc., 2017). While the data may be calculated in a variety of ways, most of the reports fall under two general categories of value-added modeling: multivariate response models and univariate response models (SAS Institute, Inc., 2016b).

The multivariate response model (MRM) is used for tests given in consecutive years (SAS Institute, Inc., 2016b). For North Carolina, this would include end-of-grade (EOG) assessments for reading and mathematics in grades 3-8. For this model, the EVAAS uses normal

curve equivalents (NCEs) to estimate gains students make from year to year. The use of NCEs is necessary when scores from consecutive assessments are not on the same scale. This MRM will take students' gains from across the state and place them on a scale similar to percentile rankings, except for the fact that with NCEs, the change from a score of 50-60 NCEs is the same as the change from a score of 80-90 (SAS Institute, Inc., 2016b). Once these scores have been converted to a common scale, the data then go through a detailed analysis to determine school and district effect sizes. It is important to note that when the EVAAS calculates teacher effect sizes it uses "a more conservative statistical process to lessen the likelihood of misclassifying teachers" (SAS Institute, Inc., 2016b, p. 10).

The second category, the univariate response model, is used for assessments that are taken in non-consecutive years (SAS Institute, Inc., 2016b). In North Carolina those assessments are 5<sup>th</sup> and 8<sup>th</sup> grade science, along with the high school course exams for Biology, English II, and Math I. This model can also be described as a "linear mixed model" and an "analysis of covariance (ANCOVA)" (SAS Institute, Inc., 2016b, p. 12). In this model, the EVAAS makes predictions based on each student's prior testing history. Every assessment that students have taken is used when calculating this predicted score, regardless of whether those assessments are in a similar subject matter. If students meet their predicted scores, their growth would be calculated as zero (0). If students exceed their predicted scores, they receive a positive growth score (a growth score above zero). If they score below their predicted scores, they receive a negative growth score (a growth score below zero). Predictions are only made for students who have at least three prior test scores (SAS Institute, Inc., 2016b).

The results of these models are used to rate teachers in North Carolina at one of three levels: does not meet expected growth, meets expected growth, or exceeds expected growth

(SAS Institute, Inc., 2017). Principals and LEA administrators can then use these results in conjunction with NCEES ratings of teachers to make decisions about a teacher's tenure in the LEA (North Carolina State Board of Education, 2016). Researchers have cautioned against using ratings such as these to make decisions regarding teacher tenure (Amrein-Beardsley, 2008; Amrein-Beardsley & Collins, 2012; Sass et al., 2014; AERA Council, 2015; Ballou & Springer, 2015). William Sanders, who is considered the founder of EVAAS, continues to defend the use of these models in the decision-making processes of schools (Viadero, 2008). Other researchers, including Cohen et al. (2003) and Hill et al. (2010) concur with Sanders.

This section provided information about how teacher evaluations have progressed since the 1700s in the United States. More detailed information was provided about the recent subjective evaluations used in North Carolina. A brief discussion of current policy was given to offer context for the usage of more objective measures in teacher evaluations such as VAMs. This discussion offered contextual information for why EVAAS ratings are used in teacher evaluations in North Carolina.

### **Theoretical Framework**

This section addresses theories that provide a foundation for assumptions regarding discrepancies that may exist between subjective and objective teacher evaluations in the state of North Carolina. Analysis of data related to these constructs could provide information to teachers and administrators about where to focus professional development for staff members that will impact student achievement. This professional development should revolve around learning theories that have become prevalent in recent years.

Social network theory provides insight into considerations to keep in mind when using subjective teacher evaluations, such as the NCEES. Social network theory posits that when

individuals believe that they are working together as a collaborative, cohesive unit, they are more likely to interact with one another and learn from each other's experiences. A principal's evaluation of a teacher using a subjective evaluation like the NCEES can create a barrier between the teacher and the principal. They are no longer seen as colleagues, but rather the positional power is exerted to a point that the teacher in question may not be as receptive to the feedback being given (Liou et al., 2015). To avoid teachers having a negative opinion about their observations, it is possible that principals could see a need to embellish teacher ratings in order to have the evaluation process be received in a more positive manner. Objective evaluations of teachers such as VAMs could provide an opportunity for principals to offer feedback to teachers based on a third-party, or neutral source. Application of this theory to the present study's questions suggests that if objective evaluations are used, then the teacher and principal will be more likely to see themselves as working together towards a common goal when providing feedback to one another.

Another theory that undergirded this study is cognitive evaluation theory. Cognitive evaluation theory asserts that if employees are intrinsically motivated by their occupation, they are more likely to be successful in that occupation. Positive feedback has been shown to increase intrinsic motivation among employees, and thus enhance job performance (Deci et al., 1999). Within the NCEES, positive feedback can be provided to the teacher in the comments section of each standard. Deci et al. (1999) found, however, that intrinsic motivation is increased more when positive feedback is received without the employee realizing he/she is being evaluated. This theory would suggest that evaluations of teachers be conducted in a way that the evaluator could be viewed more as a coach, and less as an evaluator. Using objective measures of teacher

performance could offer opportunities for principals and teachers to have conversations around ways to improve results through positive feedback.

Expertise theory should also be taken into consideration when determining actions that will contribute to improved instructional practice, as well as student achievement. Expertise theory evolved out of a study conducted by Ericsson et al. (1993). The researchers found that those who were experts in their field progressed along a continuum from novice to expert as they continued to practice in their respective roles. The skills that they possessed and needed to progress to the next level varied along the continuum. The expertise theory that arises from this research suggests that principals may need to focus on different characteristics when evaluating a teacher based on the level at which they are currently performing. Beginning teachers will exhibit certain characteristics that are very different from lead teachers. To progress to an accomplished, or expert teacher, beginning teachers will need varied feedback as they progress along the continuum. This theory suggests that principals should be able to use some discretion when deciding which teachers to evaluate and how often. Objective results of teacher performance could provide beneficial information to principals when making these decisions.

Teaching is a complex occupation. While it may be difficult to evaluate this occupation in a precise manner, evaluation is both legislated and necessary since the education system is publicly funded (Danielson, 2016). For teachers to improve on their practice, the feedback given should be precise, no matter whether it is given in a subjective, or objective manner.

These three constructs (social network theory, cognitive evaluation theory, and expertise theory) provided support for this study. They provided me with a foundation on which I created my research questions and instrumentation items. When developing the interview protocol for this study, these theories were also taken into consideration.



## **Pertinent Research and Expert Perspectives**

This dissertation sought to determine the degree to which subjective and objective teacher evaluation ratings are correlated in the state of North Carolina. Subjective teacher evaluations, such as those developed through the North Carolina Educator Evaluation System (NCEES), can have varied degrees of implementation quality and impartiality (Woulfin et al., 2015). Objective teacher evaluations, some of which employ statistical metrics such as the Educational Value-Added Assessment System (EVAAS), could overlook certain elements of schooling that have a significant impact on student performance (Karl et al., 2013). Each evaluation type has elements that could be beneficial for educators, but also areas of caution for those administering and interpreting these instruments. It is important to ensure that proper evaluations are implemented. Public schools are funded by public tax dollars, and it is the right of the citizens to have transparency in how these funds are being spent, and to know if said funding is having an impact on student achievement (Danielson, 2016). The research that follows provides information about both subjective evaluations like the NCEES, and objective evaluations like EVAAS.

### **Objective Evaluation and Value-Added Modeling (VAMs)**

A variety of sources that address value-added modeling (VAMs) and its use in schools can be found. Harris (2011) suggests that literature exists that is staunchly against the use of VAMs. Duckor et al. (2014) are themselves strong critiques of VAMs. Harris and Sass (2014) and Hill et al. (2010) believe that VAMs should be used with caution, while Wright et al. (2010) avidly support the use of VAMs. This section reviews this literature and related expert perspectives from the field of education.

The conversation about using VAMs in education dates back as early as the 1970s but became much more widespread when Bill Sanders started his work in Tennessee developing the

Tennessee Value-Added Assessment System (Stewart, 2006). Today, VAMs are used frequently when researchers measure impacts of a variety of initiatives. Such evaluation measures are part of the conversations about the effectiveness of schools, principals, and teachers (Koedel et al., 2015), teacher preparation programs (Lincove et al., 2014), and school level initiatives (Butcher & Visher, 2013). VAMs are even used to review the impacts of physical education programs (Hushman & Hushman, 2015). While Schochet and Chiang (2013) claim that these measures are unreliable, VAMs are likely to be used in public education for the foreseeable future (Stewart, 2006).

Controversies associated with value-added models have been profiled in popular literature, professional publications, and peer-reviewed studies. The conversation about VAMs has been particularly divisive when these measures are used to evaluate individual teachers (Koedel et al., 2015). Sparks (2012) contends that some researchers suggest that VAMs can provide a more objective measure of teacher performance, but also points out that others are concerned that VAMs typically only evaluate reading and math teachers which could cause other subject areas to be ignored. Darling-Hammond (2012) describes several instances of teachers being fearful of losing their jobs, where these fears are based solely on the outcomes of the VAMs of their respective states. She gives one specific example of a teacher in New York whose name was published alongside her value-added test score ratings. This teacher and her family were then harassed by the media for not being dedicated enough to the profession (Darling-Hammond, 2012). Nick Anderson (2013) reported in *The Washington Post* that 44 teachers in Washington, D.C. were given incorrect value-added performance evaluations due to an error in the value-added system used by Washington, D.C. schools (Anderson, 2013). One of those teachers was fired for the rating he/she received. The school system later attempted to reinstate

the employee when they learned that the error had occurred. Scenarios such as the ones described in this section provide validation for critiques of VAMs.

One of the key components in the Race to the Top grant that was offered by the United States' government was that states use accountability measures for teachers that included student outcomes, like EVAAS (U.S. Department of Education, 2010). One of the concerns that is present in the literature is that students in classrooms across the nation are not usually randomly assigned to their teachers (Paufler & Amrein-Beardsley, 2014). In many schools, students are assigned to classrooms based on a variety of factors. These factors often include students' current academic performance (Guarino et al., 2015). When students are tracked in such a manner and decisions about their most appropriate placements are made by principals and other officials at the school, these assignments are not considered random; thus, the results of the value-added measures of these student groups cannot be considered unbiased (Paufler & Amrein-Beardsley, 2014).

Another element of value-added modeling that is addressed by critics is the stability of the data relating to teacher effectiveness measures. Teacher effectiveness ratings often fluctuate between years, and even between student groups (Newton et al., 2010). There have been instances in which teachers receive a high value-added score one year, and in the next, or subsequent years, their scores have been substantially lower on the value-added scale (Baker et al., 2010). These fluctuations have, at times, been so drastic, that the value-added estimates for teachers had a stronger correlation with a student's test scores the previous year, than the year in which the value-added measures were applied. In one study, the value-added scores of fifth-grade teachers had a stronger correlation with students' fourth-grade achievement scores than their fifth-grade scores (Baker et al., 2010).

Proponents of VAMs suggest that measures that are more objective than those in teacher observation protocols are necessary when administrators evaluate teachers. This is because subjective principal evaluations of teachers are often swayed by the evaluator's personal opinions of those whom they are rating (Harris & Sass, 2009). Outside of the ability to provide summative teacher evaluations, however, proponents of VAMs like EVAAS assert that their utility lies in the ability of administrators and teachers to use the information gained from these measurements to elicit formative changes at the school and classroom level (Sanders & Horn, 1998). It would be difficult for such decisions to be made at the school level without the longitudinal data provided by a system such as EVAAS, or other VAMs (Sanders, 1998).

Many critiques claim that a major shortcoming of VAMs like EVAAS is that they do not account for factors such as socio-economic status (SES) and student background that impact student achievement. However, there are many factors that are related to these variables so that even if they are accounted for, the causes of the variance that may be produced are indistinguishable from these related factors (Ballou et al., 2004). For example, schools that serve students in areas of high poverty may find it more difficult to hire and retain qualified teachers for a variety of reasons. If additional controls are placed into a value-added model like EVAAS for SES, it would be impossible to determine if the effects from living in SES are due living in poverty, or the lack of a more qualified teacher. For this reason, it is unnecessary to account for these controls without being able to accurately attribute the effects to the correct causal attribute (Ballou et al., 2004).

Amrein-Beardsley and Collins present a unique perspective on VAMs. They were involved in 2012 in researching EVAAS reporting about a court case where litigation had been filed against the Houston Independent School District (HISD). Amrein-Beardsley was also called

as an expert witness in this case. She and Collins found that the way in which EVAAS results were being used as a criterion for termination, as well as for incentive pay, did not completely align with the teachers' scores. In one specific example, "Teacher A" received a higher teacher effectiveness score for math than she/he had in the previous years and was still terminated. In the previous three years, even though her/his EVAAS scores were lower, Teacher A received a bonus from the incentive pay program and received no supplemental pay in the year she/he was terminated. Such findings were originally identified by Amrein-Beardsley in a 2008 study (Amrein-Beardsley, 2008). At that time, she suggested that VAMs could be beneficial when used in education but cautioned against using them as part of teacher evaluation processes. Collins engaged in research of her own in 2014 and found that teachers were averse to how EVAAS was being used in their systems. Collins also cautioned against using EVAAS results in a high-stakes manner.

Condie et al. (2014) also stated that VAMs can be beneficial in terms of discerning student gains in academic achievement, but that they should not be used to evaluate teachers. Darling-Hammond et al. (2012) advocate for the use of VAMs in research, but again cautioned against using them in teacher evaluations. Franco and Seidel (2012) concluded that, based on its use in Ohio, EVAAS can provide beneficial information to educators. However, they also asserted that there are circumstances that take place within a school, especially a school serving high proportions of students in poverty, for which EVAAS cannot adequately account. These factors could alter student achievement scores. Marchant et al. (2015) cautioned against using student outcomes as a part of teacher evaluations altogether. They pointed, instead, to using the Interstate Teacher Assessment and Support Consortium (InTASC) Standards for teacher evaluations. Many of the InTASC Standards align with rubrics created by Danielson and

Marzano for teacher evaluations. Sass et al. (2014) insist that VAMs cannot be unbiased in their usage, and thus, will always be skewed.

In this section, I provided insight into varying perspectives of VAMs. Controversial scenarios were presented from recent literature to explain the concern over using VAMs as a sole measure for rewarding effective teachers or pointing out ineffective teachers. Several perspectives were presented from researchers that suggested using VAMs with caution and in conjunction with subjective measures of teacher evaluation.

### **Subjective Evaluation Measures**

Extensive research exists for subjective teacher evaluations, with divisions of support similar to those associated with objective measures. Baker et al. (2010) are strong supporters of subjective teacher evaluations. Darling-Hammond et al. (2012) advocate for the use of such standards as those provided by the Interstate New Teacher Assessment and Support Consortium (INTASC) in conjunction with VAMs. Rockoff and Speroni (2010) also advocate for teacher evaluations that use both subjective and objective measures. On the other hand, there are individuals who have opinions similar to that of Hartford Superior Court Judge Thomas Moukawsher who compares such subjective measures to “cotton candy in a rainstorm” (Thomas, 2016). To Judge Moukawsher’s point, principals have been perceived by their superintendents to have difficulty in effectively carrying out subjective evaluations to make decisions about teacher dismissals (Ward, 1995).

Subjective evaluations have been shown to be useful provided that certain measures of fidelity are taken into consideration. Gargani and Strong (2014) argue that subjective evaluations like the NCEES are far too long and arduous to provide beneficial feedback and recommend the use of the Rapid Assessment of Teacher Effectiveness (RATE), which is also a subjective

evaluation, but one that they believe is much more streamlined and efficient. Based on their evaluation of the state of Massachusetts' implementation of The Model System for Educator Evaluation, an evaluation system similar to the NCEES, Dowley and Kaplan (2014) suggested that North Carolina administrators proceed with caution. The authors found that this evaluation model could be used effectively, but that the effectiveness of the implementation varied across the state. Taylor and Tyler (2012) contend that teachers tend to perform better in terms of student achievement during a year in which they are being evaluated, and that this improvement carries into subsequent years following the evaluation year. With this rationale, it seems plausible that evaluating teachers causes them to become more cognizant of their classroom and improves their practice.

A benefit of subjective evaluations is their ability to be used as a tool to provide specific feedback to teachers that would encourage changes to professional practice (Templeton et al., 2016). For such change to come about and have lasting impacts on student achievement, the subjective evaluations provided by school administrators should offer extremely detailed feedback so that educators are aware of how to improve their practice (Taylor & Tyler, 2012). Such detailed feedback can provide benefit not only for the teacher being observed, but also for the observer. An important dimension of a useful and effective subjective evaluation is when that evaluation system improves the educational practices of those in the role of observer, so that that benefit is then passed on to all teachers with whom they interact (Archer et al., 2016).

While the subjective evaluations described in the previous paragraph may be considered optimal, some critiques point out that there is often little variation among teacher ratings when administrators employ subjective evaluation measures (Weisberg et al., 2009). Weisberg et al. (2009) have coined this lack of variation as the "Widget Effect," a term that refers to "the

tendency of school districts to assume classroom effectiveness is the same from teacher to teacher” (p. 4). This Widget Effect may exist because principals have a variety of reasons to inflate teachers’ evaluations (Danielson, 2016). Social network theory would suggest that principals may be attempting to stay within the realm of being associated with teachers as colleagues instead of as evaluators for fear that their feedback may not be received in a positive manner. Cognitive evaluation theory would assert that if the feedback given to teachers is not received in a positive manner, then teachers are likely to experience a reduction in morale and be less productive. Regardless of whether these evaluations are exaggerated, many subjective evaluations do not provide useful feedback to those being evaluated to begin with (Harris et al., 2014). With the teacher shortage that is present in the United States (Yaffe, 2016), it is reasonable to consider that principals may rate teachers at a higher level than they consider them to be performing in order to avoid having to try to replace them.

Charlotte Danielson has similar opinions of the subjective evaluations that exist today. Danielson supports subjective evaluations like the one found in her work *Enhancing Professional Practice: A Framework for Teaching* (1996), but believed that there are too many incentives for principals to “inflate teacher ratings.” One such motivation is the sheer amount of time that it takes to conduct the subjective evaluations, as well as the additional responsibilities districts place on principals if a teacher rated poorly on this measure, such as improvement plans (Drake et al., 2015; Neumerski et al. 2018). Danielson thinks that the subjective measures that are used today require far too much on the part of principals, and far too little participation from teachers. This practice does not yield very many benefits with regards to improving teaching practice. What would be more beneficial would be a process in which teachers collaborate with



their colleagues to solve various problems. This has greater potential to improve how teachers teach (Danielson, 2016).

### **Debate over the Necessity of Both Subjective and Objective Evaluations**

This section and subsections presented both critiques and supportive arguments for both subjective and objective teacher evaluations. Also present in the literature are perspectives of researchers that assert that both forms of evaluation can provide benefits to improved teaching practice and increased student achievement when used in conjunction with one another. This section will present these perspectives as they exist in current literature. Reasons for discrepancies in subjective and objective teacher evaluations could be more rational than one might expect. Subjective teacher evaluations often encompass a variety of skills, including factors that may not be directly related to an increase in student achievement; however, these skills may still provide benefit to the functioning of the school (Harris et al., 2014). Such skills that principals may consider to be important are how caring the teacher is and whether they participate in school activities outside of the classroom (Harris et al., 2014). Teachers may be rated as effective in these areas, but the ratings may have little correlation with any changes in student achievement. The opposite could also be true of teachers who focus solely on making sure that students do well on their standardized assessments. Some teachers who have high value-added scores are viewed as “lone wolves” (p. 101) by their principals (Harris et al., 2014). These teachers would receive lower ratings on subjective measures where teachers are evaluated based on their willingness to collaborate with their colleagues. The possibility arises from this discussion that these two evaluative methods may be designed to measure different teacher qualities. A lack of correlation would then be expected.

Since teaching is an exceedingly complex occupation that requires a specific skillset that is not obligatory in other professions (Grossman et al., 2009), it would be unwise to rely on a single measure of teacher effectiveness as the sole source of information when making decisions regarding pay or tenure (Hinchey, 2010). For teacher evaluations to impact student achievement, they should be based on factors that are highly correlated with increased student achievement, as well as the continual growth of the educators being evaluated (Hinchey, 2010). When subjective evaluations are focused on the growth of the educator, and multiple evaluators provide feedback to the teachers being evaluated, the correlation to value-added estimates of student achievement can be relatively strong (Heneman III et al., 2006). This dissertation will seek to determine if similar correlations exist between the subjective teacher ratings in the NCEES and the objective ratings within the Teacher Effectiveness Reports in EVAAS.

### **Chapter Summary**

In this chapter, I presented expert perspectives pertaining to both subjective and objective teacher evaluations. Many of these perspectives offered views that were in strong opposition to each other. While subjective evaluations such as the NCEES and objective measures such as EVAAS can provide for interesting topics of debate, there seems to be a consensus around the idea that teachers should be evaluated using multiple measures when high-stakes circumstances are involved (Rice et al, 2012). Researcher perspectives about using both teacher evaluation methods in conjunction with each other were also presented in this chapter. North Carolina is currently using both subjective and objective teacher evaluations in the form of NCEES and EVAAS, respectively. Chapter 3 of this dissertation discusses the methods that were employed to determine if a relationship exists between these two evaluation systems in North Carolina, as well as the perspective of principals on these measures.

## **CHAPTER 3: METHODOLOGY**

### **Chapter Introduction**

The purpose of this study was to gain insight into the perceptions of principals in the Northeast and North Central regions of North Carolina regarding the correlations between EVAAS and NCEES ratings of middle grades mathematics teachers. Principals who agreed to participate in this study were given access to the most recent correlations of EVAAS and NCEES ratings in their school district and were asked to respond to interview questions related to these correlations. The Interview Protocol can be found in Appendix A.

In this chapter, I provide details about how the research study was conducted. A brief description of both EVAAS and NCEES measures of teacher evaluations are offered, as well as specifics about how these metrics are used in North Carolina. The participants, data collection methods, and data analysis procedures are also described here. I chose to focus specifically on principal opinions on EVAAS and NCEES correlations of middle grades mathematics teachers due to research that concludes that mathematics achievement is particularly impacted by instructional factors within the school (Marzano, 2003; Parcel & Dufur, 2001).

The qualitative elements of the study aligned with the description of a phenomenological study provided by Creswell (Creswell, 2014). The phenomenon studied was the correlation between principal ratings of teachers on the North Carolina Educator Evaluation System (NCEES) and the Education Value-Added Assessment System (EVAAS) ratings of these same teachers. In recent years, these two data sets have proven to be very weakly correlated (Batten et al., 2012; Henry & Guthrie, 2015). The study examined archival data in the most recent sets of NCEES and EVAAS ratings to determine the level of correlation between these constructs.

These correlations were the central focus for the phenomenological elements of this study, which explored principals' perceptions of the level of correlation between the scores.

### **Research Design**

For this research study, I used a qualitative interview protocol. I received district level NCEES and EVASS correlations of teacher ratings from the superintendents in participating districts. Those superintendents were provided with these data by the North Carolina Department of Public Instruction (NCDPI). I emailed an informed consent document (Appendix C) to superintendents of school districts in the Northeast and North Central regions of North Carolina to ask for permission to interview middle school principals. In districts in which superintendents provided consent, I emailed principals directly to request their participation in an interview regarding these correlations. For one specific district, there was a request for me to not email principals directly, but instead to send a general email to a representative from the district who would then communicate with the principals in that district. I complied with this request.

A qualitative interview protocol was chosen for this study so that I could gain specific input from principals about their opinions regarding 1) how the correlations of teachers' NCEES and EVAAS ratings could be impacted by a variety of factors, 2) whether they believed that these correlations aligned with their expectations based on contemporary data, and 3) the degree to which they believed that it is important for EVAAS and NCEES ratings of teachers to be correlated. By interviewing the participants, I provided them with an opportunity to elaborate on their understanding of both the NCEES and EVAAS ratings and what these ratings measure. The interviews were conducted via Zoom due to COVID-related restrictions.

A phenomenological study was chosen so that participants could describe their lived experiences with regards to EVAAS and NCEES correlations. Conducting this research in this

way allowed me to understand what this phenomenon meant to each individual participant, and Johnson and Christensen (2000) noted this capacity as a strength of the phenomenological approach. This allowed the complexities with which each participant engaged with these correlations to be revealed without any preconceived notions; Husserl (2012) describes this as an additional utility of this research approach. The phenomenological process benefited the study by allowing participants the opportunity to elaborate on their experiences openly and their responses were reviewed to find themes that span across each of their individual experiences.

### **Research Questions and Hypotheses**

Relationships among the study's constructs were addressed through the following research questions:

1. How do the correlations of NCEES and EVAAS ratings compare among the participating districts in the Northeast and North Central regions of North Carolina?
2. How do principals describe the levels of correlation that exist between the district's NCEES and EVAAS ratings?
3. How do principals describe the levels of correlation that exist between their teachers' NCEES ratings and EVAAS ratings and those of the district?
4. To what factors do principals in the Northeast and North Central regions of North Carolina attribute the level of correlation between the NCEES and EVAAS ratings of teachers?
5. What teacher attributes do principals in the Northeast and North Central Regions of North Carolina believe would increase student achievement and thus produce a stronger correlation in NCEES and EVAAS ratings, if an evaluation of these attributes existed in the NCEES?

## Study Participants

### Participant Descriptions

I conducted this study after receiving approval from the North Carolina State University Institutional Review Board (IRB) and from superintendents in selected districts. The interviews were conducted with principals in the Northeast and North Central regions of North Carolina. Six out of seven of the principals who participated came from schools who served a high population of students who came from an economically disadvantaged background; i.e., at least 40% of the student body qualified for free or reduced priced lunches. The ratio of 40% was chosen to align with the requirements for a school to be considered for Title I schoolwide programs. I required that each principal participant had been a principal at her/his site for at least one year to participate in the interview process. While background and demographic data were collected on each participating principal, their respective identities were kept confidential. I conducted convenience sampling based on my proximity to the regions selected. In total 35 superintendents were contacted to request their participation in the study, 5 districts agreed to participate, but only 4 districts had principals who actually replied to my request to be interviewed. Within those 4 districts, 7 principals agreed to participate in the interview process.

Once participating principals were identified, I sent the names and email addresses of the participating principals to a representative at the North Carolina Department of Public Instruction (NCDPI), who provided to the superintendent of each participating district the correlations between their NCEES ratings of teachers and the EVAAS ratings of their middle grades mathematics teachers. This information was used when the principals were responding to the interview questions (Appendix A). I emailed principals directly in districts whose superintendents consented to the study (aside from one district that asked me to email a

representative who would then reach out to the principals) with the intent of locating at least 6-8 principals who would agree to be interviewed for this study. Pseudonyms, both during the interview and in the reporting of results, were used for the participating principals.

In addition to EVAAS and NCEES correlations data derived from the above-mentioned participants, this study employed information about the demographics of participating schools and districts secured from the North Carolina School Report Cards. Six out of seven of the principals who were interviewed for this study came from schools that serve a population in which the majority is constituted of students who are low-income. This determination was based on the percent of students who qualify for free or reduced-price lunches. Of the principals who were interviewed, 50% or more also served schools that received a D or F rating on the North Carolina school accountability rating system during the 2018-2019 school year. One principal served in a school in which less than 40% of students qualified for free/reduced price lunches.

The focus on middle grades was determined due to research that focuses on middle grades mathematics teachers. These studies conclude that teachers have a significant impact on student learning in mathematics (Chiang et al., 2017; Ye & Singh, 2017). This is in contrast with research that suggests that there could be multiple outside factors that more significantly affect student learning in reading (Harlaar et al., 2007; Jennings et al., 2013). Reading and mathematics were the main areas that I considered when determining the focus for this study, since EVAAS ratings are provided for teachers in many grade levels for teachers of both of these subjects. Science teachers only receive EVAAS ratings if they are in grades 5 or 8, or if they teach Biology.

Subjects for which teachers receive EVAAS ratings were important to this study due to North Carolina's legislation in recent years that utilized EVAAS scores. During the 2014-2015

and 2015-2016 school years, EVAAS scores were used as a stand-alone measure of teacher and administrator effectiveness. These ratings were used as Standard 6 of a the state rubric which served as the basis of a teacher's evaluation, and Standard 8 of the rubric for an administrator's evaluation (Davis et al., 2015). Subsequent to this two-year period, the State of North Carolina removed Standard 6 from teachers' evaluations, and Standard 8 from administrators' evaluations (North Carolina State Board of Education, 2016). EVAAS reports may still be used when determining a teacher or administrator's ratings, but they are no longer used as a stand-alone measure (North Carolina State Board of Education, 2016). The aggregate ratings of all teachers who taught a subject for which an End-of-Grade (EOG) or End-of-Course (EOC) assessment was administered still impact the overall School Performance Grade (SPG) for each school. In the 2018-2019 school year, 20% of a school's SPG came from the EVAAS growth measures, and 80% of a school's SPG came from the overall proficiency ratings of the students in the school (North Carolina School Report Cards, 2018).

### **Variables**

For the quantitative element of this study, I reported archival data and compared these results to the district specific data that were provided by the NCDPI. These data were initially sent to the superintendents of participating districts and then sent to me by the superintendents. For the qualitative element of this study, I focused on principal perceptions of the NCEES and EVAAS ratings of middle grades mathematics teachers in their district and their analysis of the correlation of these two evaluative measures. Descriptive statistics were also provided for these two variables. Demographic data for the participating principals' schools were also provided to give an accurate depiction of the schools from which the data were drawn. This information was



gathered from the North Carolina School Report Cards (2018), as well as responses from participants.

The calculations that are used in EVAAS to determine Teacher Effectiveness ratings are an important element to address. SAS Institute, Inc. (2016b) produced a whitepaper that gives some description of how these ratings are generated. For middle grades mathematics teachers, the model that is used in EVAAS is a multivariate response model (MRM). This model can also be described as a “linear mixed model,” or “multivariate, repeated-measures model” (p. 3).

According to this document:

The MRM is a gain-based model, which means that it measures growth between two points in time for a group of students. The growth expectation is met when a cohort of students from grade-to-grade maintains the same relative position with respect to statewide student achievement in that year for a specific subject and grade. The key advantages of the MRM approach can be summarized as follows:

- All students with valid data are included in the analyses, even if they have missing test scores. All of each student’s testing history is included without imputing any test scores.
- By including all students in the analyses, even those with a sporadic testing history, it provides the most realistic estimate of achievement available.
- It minimizes the influence of measurement error inherent in academic assessments by using multiple data points of student test history.
- It allows educators to benefit from all tests, even when tests are on differing scales.
- It accommodates teaching scenarios where more than one teacher has responsibility for a student’s learning in a specific subject/grade/year.

- The model analyzes all consecutive grade subjects simultaneously to improve precision and reliability. (SAS Institute, Inc., 2016b, pp. 3-4)

This MRM is described as a “gain-based model” (p. 3) because it “measures growth between two points in time for a group of students” (p. 3). To determine this growth, EVAAS converts student scale scores to normal curve equivalents (NCEs) so that student growth can be measured from year to year regardless of which assessment they are taking (SAS Institute, Inc., 2016b). NCEs are scaled in a manner that appears to be similar to percentiles. The mean, or average, NCE score for students will, by definition, be 50. Those students who score above 50 NCEs are deemed to have made some growth beyond the norm. Students who score below 50 NCEs are considered to have made less growth than the average student in North Carolina for that particular school year (SAS Institute, Inc., 2016b). Middle grades mathematics teacher ratings in EVAAS are determined based upon the NCE scores of all students who are assigned to them. Those scores are then compared to the scores of other teachers who have students with similar testing histories to determine each teacher’s effectiveness (SAS Institute, Inc., 2016b). The estimates used for teacher ratings are more conservative than those used for district and school models “to lessen the likelihood of misclassifying teachers” (SAS Institute, Inc., 2016b, p. 10).

Teachers in North Carolina also receive ratings that would be considered subjective according to the definition of researchers Jacob and Lefgren (2008). Those ratings are based on how the teachers are scored on the rubric within NCEES (McREL, 2015). The rubric is used by principals and their designee(s) (henceforth principal) to evaluate teachers on a scale from not demonstrated to distinguished. These evaluations can be considered formal or informal depending on the evaluation cycle for a given teacher. Formal observations must last for 45

minutes, or an entire class period. Informal observations must only last 20 minutes (McREL, 2015). Teachers can be placed on one of three evaluation cycles depending on their years of experience. Teachers who have less than three years of experience are placed on the comprehensive cycle. This cycle consists of three formal observations from the principal, and one from a peer. On this cycle, teachers are rated on all five standards. Teachers who have greater than three years of experience are usually placed on either the standard or abbreviated cycle, although they can be placed on the comprehensive cycle if the principal deems it necessary (McREL, 2015). If a teacher is on the standard cycle, she/he will receive three observations, the first of which will be formal, and the other two can be formal or informal. Teachers on this cycle will also be rated on all five standards (McREL, 2015). A teacher on the abbreviated cycle is only required to receive two observations throughout the school year from the principal, and both can be formal or informal. On this cycle, teachers are only rated on standards 1 and 4 (McREL, 2015). On all three cycles, teachers can receive additional observations if the principal deems it necessary (McREL, 2015). Such variation in the number of observations teachers receive before being provided an overall rating at the end of the year could also impact the correlations between NCEES and EVAAS ratings.

The primary variable for the study is principal perceptions of the factors that impact the level of correlation between their ratings of teachers on the NCEES standards and the EVAAS scores for these same teachers. Recent studies have shown that there could be a variety of reasons to explain the variation between subjective and objective measures of teacher performance. In some instances, teachers who are rated highly on such VAMs, such as EVAAS, may be primarily focused on what goes on in their classroom, yet part of their subjective ratings encompass how involved they are with the school and community (Harris et al., 2014). It is also

important to remember that subjective evaluations of teachers often rate educators on variables that may not show up in VAMs. These variables may include teacher enthusiasm, interpersonal skills, and their ability to work with other teachers (Harris & Sass, 2014). Such skills may not directly impact student achievement scores but may impact classroom culture.

### **Instrumentation**

The quantitative elements of the research design did not require a formal instrument for procurement of data. The NCDPI was able to provide the district level correlations of EVAAS and NCEES measures of middle grades mathematics teachers. These data were deidentified before they were sent to me. The only identification on these data was the name of the school district. The demographic information for each school was acquired during the interview process and then entered into a spreadsheet.

For the qualitative elements contained in this study, I designed an original interview protocol, which can be found in Appendix A. The purpose of the items in the protocol was to glean the perceptions of principals regarding correlations between EVAAS and NCEES ratings. With this in mind, the questions addressed the teacher attributes that principals believe contribute to an increase in student achievement, and whether such attributes are evaluated within the NCEES rubric. Historically, the correlations between EVAAS and NCEES ratings have been low, so I sought principal input on factors that may cause such a discrepancy. I began by asking about participants' background information, including information about their career as a principal (e.g., For how long have you been a principal?; For how many years have you worked in this district?; For how many years have you worked in this school?). From there, I proceeded to asking participants questions about EVAAS and NCEES correlations at the state level, and then district level. Questions were broad, since this study was phenomenological in design; such

questions are designed so as not to lead participants in any specific direction with their reply (Creswell, 2014). From these questions, the goal was for participants to describe their experiences with both evaluation systems in order to provide context for district level results, which traditionally show a weak correlation between the two systems. Questions were asked in such a way that participants understood that I was asking about their individual experience; the combination of their response enabled me to make connections to the phenomenon on a more global scale.

I convened a panel of experts to validate the interview instruments. The panel of experts included a former state superintendent, the director of human resources for a state education agency, an associate superintendent of human resources for a local North Carolina school district, and an assistant professor of education leadership and policy who has a research agenda that is related to my study topic. These experts were asked to review the degree to which the interview questions could be readily understood in order to facilitate knowledgeable responses from participants, and secure data that was useful in addressing the research topics. The panel was also asked to complete a validity questionnaire to provide specific feedback on the interview protocol (Appendix B).

Principals were asked to respond to the interview questions via an online teleconferencing platform due to COVID-19 restrictions, and I recorded these responses using both hand-written notes and an audio recorder. These interviews were then transcribed using Rev.com. Once these interviews were transcribed, I began to code the qualitative data using an in vivo coding system and then started to determine themes that emerged from the words and phrases that participants used in their responses. I did not have a preset list of words or phrases

that I was looking for when I reviewed their responses: their responses revealed the words, phrases, and themes that were important as I analyzed these data.

### **Procedures**

I secured approval from my dissertation committee for this protocol. I then received approval for this protocol from the Institutional Review Board (IRB) of North Carolina State University. After approval was granted, I then emailed superintendents from the Northeast and North Central regions of North Carolina to request permission to contact principals within their district. Once permission was granted by the superintendent, I then emailed principals to request their participation in the interview process. Six of the seven interviewees came from Title I schools that received a D or F rating for the 2018-2019 school year. Participants were emailed an informed consent document (Appendix D) and were asked to provide a signed copy to me before participating in the interview. Six out of seven of the interviewees came from Title I schools. Participants were emailed an informed consent document (Appendix D) and were asked to provide a signed copy to me before participating in the interview.

### **Accessing Data**

This study focused on principal perceptions of EVAAS and NCEES correlations in the Northeast and North Central regions of North Carolina. State level correlations of these data have been reported by the NCDPI in recent years. To gain access to these archival data sets that address the NCEES and EVAAS ratings of teachers, I emailed the NCDPI to formally submit a data request according to their guidelines and policies. Once approved, these data were sent to the superintendents of participating districts, and the superintendents then forwarded those data to me. To locate participants for the study, I also emailed superintendents to gain permission to interview principals in their districts, and then local principals to serve as actual participants.

Once I located the participants, I scheduled the interviews with principals who agreed to participate within a 12-week period. District level correlations of EVAAS and NCEES ratings were requested from the NCDPI for the districts where participants were employed.

### **Data Collection**

After receiving the quantitative data from the NCDPI, I reviewed the correlational analysis that was provided, as well as reported the descriptive statistics on each standard for both NCEES and EVAAS evaluations. These data were used to provide further insights into the evaluation process for teachers in North Carolina.

I interviewed principals who agreed to participate in the study using the interview protocol located in Appendix A and described in the previous section on instrumentation. The information gained during these interviews was kept completely confidential. The interviews were scheduled at a time that enhanced the prospects of the participant feeling comfortable speaking about the topics being discussed. I also assured each participant of the confidentiality of both his/her participation and his/her responses. This allowed the participant to speak openly about his/her opinions regarding the interview questions. These interviews were recorded with a voice recorder.

When these processes were completed, the audio files were deleted so that confidentiality could be maintained. Pseudonyms were used in place of participant names when I presented the findings of the study. I took notes during the interviews regarding elements of the interview that I found to be of particular importance. I also recorded these interviews using an audio recorder. No video recordings were conducted for this study. The audio recordings were transcribed via Rev.com. I then coded the data using an in vivo coding system (Manning, 2017) to reveal

emergent themes. The files containing the transcripts, meeting notes, in vivo coding and emergent themes were stored on my personal password-protected computer in a locked office.

### **Data Analysis**

Once I received the NCEES and EVAAS ratings of middle grades mathematics teachers from the NCDPI, I assessed the correlation in these data. I first reported descriptive data results, including the median, maximum, minimum, range, interquartile ranges, and any outliers that existed in these data. These results were evident in the graphs sent by the NCDPI.

The qualitative data were gathered through participant interviews. Their responses were transcribed and reviewed. I then analyzed these response data using in vivo coding to reveal emergent themes. The themes that emerged from their responses were used to further structure the analysis and findings of this study.

While reviewing participant responses, words or phrases of particular importance that were used were coded and entered into a spreadsheet. As similar words or phrases were mentioned in other participant interviews, I recorded such instances in the spreadsheet as additional responses for which the meanings were similar. As aligned responses continued to come up in the participant interviews, themes emerged that further enabled me to categorize those responses. In total, 123 codes were used to develop the themes that are listed in the following paragraphs.

Research Question 1 (*How do the correlations of NCEES and EVAAS ratings compare among the participating districts in the Northeast and North Central regions of North Carolina?*) was quantitative in nature and thus did not require the use of in vivo coding.

For Research Question 2 (*How do principals describe the levels of correlation that exist between their district's NCEES and EVAAS ratings?*) two themes emerged from participant



responses. Those two themes were 1) there were expectations of some correlation between Standard 4 in NCEES and EVAAS ratings and 2) there were little to no correlation between NCEES and EVAAS.

For Research Question 3 (*How do principals describe the levels of correlation that exist between their district's EVAAS and NCEES ratings compared to their daily observations of teachers within their school?*) two themes emerged. Those two themes were 1) there are correlation at the extremes and 2) NCEES can be easily manipulated.

For Research Question 4 (*To what factors do principals in the Northeast and North Central regions of North Carolina attribute the level of correlation between the NCEES and EVAAS ratings of teachers?*) two themes emerged. Those themes were 1) school and/or district level interventions are not factored into teacher ratings and 2) NCEES is a very subjective rating scale.

For Research Question 5 (*What teacher attributes do principals in the Northeast and North Central regions of North Carolina believe would increase student achievement and thus produce a stronger correlation in NCEES and EVAAS ratings, if an evaluation of these attributes existed in NCEES?*) two themes emerged. Those two themes were 1) relationships and 2) pedagogy.

The themes above are addressed in detail in detail in Chapter 4. I maintained these data on a password protected device that was only accessed by me and was kept in a locked location when I did not have it with me. Participant anonymity was protected at all times throughout the duration of this study. At no point during the interviews were participants asked to give their names so pseudonyms were not required for this study.

## Subjectivity Statement

I have worked in the field of education for nearly 17 years. During this time, I have held a variety of positions, including that of a principal. Throughout my career teacher evaluations have been an a priority for me; I have been evaluated in the role of teacher and I have also frequently been responsible for assessing the performance of such practitioners. I have a particular interest in correlations between NCEES and EVAAS ratings. In addition to having experience as a principal, I also served as a data coach for a school district earlier in my career. During that time, I attended several trainings on EVAAS and spent a lot of time reviewing district, school, and teacher level data for the district that I served.

From my personal experiences with colleagues in the field and my own experiences in the previously-mentioned roles, I have preconceptions about correlations between NCEES and EVAAS measures of teacher performance. Throughout my dissertation I worked to ensure that my own perspectives did not influence the results of this study. One way that I did this was to use one of the techniques of bracketing described by Tufford and Newman (2010). I routinely conducted interviews with trusted colleagues and my committee members so that they could provide me with feedback and insight on personal bias that they may have noticed in my research. I also used theory triangulation and methodological triangulation methods as described by Bhandari (2023). My conceptual framework consisted of three theories: Social Network Theory, Cognitive Evaluation Theory, and Expertise Theory. These theories were triangulated to help ensure the researcher bias was addressed. For methodological triangulation, I used quantitative depictions of correlation data and qualitative perspectives of participants and juxtaposed these with my own expertise gleaned from significant training in both forms of evaluation. While it is likely impossible to remove all researcher bias, by providing this context

the reader can evaluate these findings in the context of my lived experiences and make her/his own judgment as it relates to my subjectivity on this topic.

### **Delimitations of the Study**

The delimitations of this study are as follows:

1. This study only used data from the 2018-2019 school year in the analysis of the relationship between ratings of middle grades mathematics teachers within the NCEES and the Teacher Effectiveness Reports produced by EVAAS.
2. The teachers whose NCEES and EVAAS scores were used in the analyses for this study were limited to middle grades teachers and the standardized scores were limited to those from mathematics assessments.
3. The locations of participants were limited to the Northeast and North Central regions of North Carolina.

### **Assumptions**

This study was guided by the following assumptions:

1. I assumed that participants responded to interview questions honestly and without fear of negative consequences for their responses.
2. I assumed that the transfer of archival data from the NCDPI to data collection sheets was accurate.
3. I assumed that participating principals understood EVAAS calculations when they responded to interview questions regarding EVAAS and NCEES ratings.
4. I assumed that participants, after I provided a preliminary briefing, understood the data depictions in the graphics provided by the NCDPI.

5. I assumed that teachers for whom we received data from the NCDPI were rated according to the NCEES guidelines with respect to the number of observations they received before being given an overall rating in NCEES.

### **Chapter Summary**

The purpose of this study was to determine the perceptions of principals in the Northeast and North Central regions of North Carolina regarding the level of correlation between NCEES and EVAAS ratings of teachers. To address the research questions presented in this dissertation, qualitative measures were used once I gained IRB approval. The participants in this study were principals from districts in the Northeast and North Central regions of North Carolina. I requested access to quantitative data from the NCDPI via email and an online data request process. I requested permission to interview principals via email, as well. Once the interviews were completed, they were transcribed using Rev.com. After the transcriptions were completed, I analyzed these data using an in vivo coding system to reveal emergent themes.

In recent years, many states moved entirely away from using student test results as part of teacher evaluations (Shapiro, 2019). Even though states started to trend towards not using such data in teacher evaluations, many use these data to adjust and improve the professional development that they offer teachers, as well as their overall evaluation systems (Hawkins, 2019). North Carolina also uses the data as a basis for awarding teacher bonuses and determining the accountability grades for schools. The use of these data in teacher evaluation, while no longer mandated, continues in this and other states. This study helped to inform the practice of teacher evaluations in the State of North Carolina in the context of an ongoing debate about the roles of student assessment data in evaluation and accountability practices.

Results of the above-mentioned analyses are reported in Chapter 4. Chapter 5 provides a discussion of the results and implications of the study for policy and practice. Recommendations for future research are also included.

## **CHAPTER 4: RESULTS**

### **Chapter Introduction**

The purpose of this study was to explore the perspectives of middle school principals who serve in the Northeast and North Central regions of North Carolina regarding the correlations between two teacher evaluation systems, the Educational Value-Added Assessment System (EVAAS) and the North Carolina Educator Evaluation System (NCEES). Such correlations have been reviewed by various researchers in the field of education, but I believed asking principals directly about their perspectives regarding such correlations within their specific district would provide valuable insight for this topic. Specifically, I wanted to see what attributes of a teacher principals believed would contribute directly to student achievement and if there would be greater correlation between NCEES and EVAAS if those attributes were addressed in NCEES.

Over the span of approximately two months, I interviewed participants via Google Meet or Zoom online meeting platforms due to COVID-19 restrictions. As I proposed, I was able to interview seven principals. These principals served four different school districts, and six of the seven principals served schools that were considered high-poverty based on the percentage of students who received free and reduced priced lunches at these schools. The participant names were never used during the interviews, so pseudonyms were not necessary when conducting the interviews. I interviewed the participants using the protocol found in Appendix A. Each interview was audio recorded and once all interviews were conducted, the recordings were transcribed using the online transcription service Rev.com. All participants were able to provide valuable insights into the questions from the protocol. In instances where the participants needed additional clarity about the matters addressed in the protocol, such insight was provided so that

they could adequately respond to the interview questions. I also provided guidance when necessary for participants who had questions about how to read and analyze the graphics that were provided by the NCDPI.

Data collected from participant interviews provided substantial insights into principal perceptions about correlations between the scores on the Educational Value-Added Assessment System (EVAAS) and ratings on the North Carolina Educator Evaluation System (NCEES). These interviews further illuminated their opinions on the teacher attributes that contribute to an increase in student achievement. The participants also provided feedback on training provided on these accountability measures, or the lack thereof.

As someone who grew up in rural eastern North Carolina, and whose entire career has been spent in eastern North Carolina and largely in schools that were classified as high-poverty, I was eager to receive feedback from other principals about the teacher evaluation process. While conducting these interviews, I was able to gain insights into the perspectives of principals as it relates to what is arguably one of the most important parts of their role, evaluating teachers. The interview process also allowed these principals the time to reflect on what attributes of their teachers they believe yield results and increase student achievement. Once participant interviews were complete, the recordings were sent to an online transcription site (Rev.com). Once the transcriptions were complete, I then coded the transcriptions based on the themes that emerged.

Data collected from the interviews that were conducted provided great insight into principal perceptions of the correlations between how teachers are rated in the North Carolina Educator Evaluation System (NCEES) and the Educator Value-Added Assessment System (EVAAS). Principals also provided insights regarding the attributes of a teacher that they believe

contribute to an increase in student achievement, and whether those attributes are adequately addressed in the NCEES.

As a principal at the time the interviews were conducted, I was interested in gaining such insights from colleagues in districts other than my own. From these interviews I was able to learn how districts prepare principals to conduct the teacher evaluation process and how much ongoing support is provided. It was also intriguing to me to hear from other principals the attributes that they believe contribute to a teacher being able to increase student achievement in the classroom. These individuals were also forthcoming about the difficulties regarding the NCEES and how that could contribute to diminished correlations between the NCEES ratings and EVAAS scores.

At the beginning of the interview process, there were background questions for each participant. These questions provided information on the years of experience each principal had, and the demographics of their school and the community in which they served. Their responses to these questions provided context for the subsequent questions in the protocol that related to the NCEES and the EVVAS. The next section provides a profile of each participant based on these questions.

### **Participant Profiles**

As was mentioned previously, pseudonyms were not necessary for this study because it was not necessary for participants to use their names to respond to the questions in the interview protocol. Participant profiles are described in the following paragraphs to provide context for the study.

Participant 1 was a White female who was entering her ninth year as an administrator at the time of the study. She had been the principal of an elementary school for six years, and she was in her third year at a middle school. Between those two administrative roles, she served as a



director in the human resources division of a school district. With this array of experience, Participant 1 was able to provide unique insights into her observations of EVAAS and NCEES correlations.

Participant 2 had been a principal for seven years. The school at which she served during the 2018-2019 school year was self-identified as a rural school. She is a White female who had participated in the initial “rollout” of district trainings related to the NCEES system. In addition to being present in the initial district meetings when NCEES was being instituted, she was also partly responsible for training other principals on the NCEES system. Her knowledge related to NCEES provided substantial context to the ratings related to her district.

Participant 3 was a White male who was in his fifth year as a principal. All of his experience was in the same school district and at the same school. This participant felt that he had received enough training on the EVAAS system that he was very familiar with the trainings that would be presented at the district level; however, he did not remember being trained at all on the NCEES. This participant did have a knowledge of correlation coefficients and was able to apply his knowledge to his responses in this interview.

Participant 4 was an African American male who had been a principal for thirteen years, and he was in his third year in both his school and district. He felt that he had rather extensive training on the NCEES in his previous district. However, he did not feel that he had much training in either district regarding the EVAAS. He did attend trainings at the regional level, but he did not feel that adequate training had been provided at the district level of his school systems.

Participant 5 was a White female who had been a principal for approximately six years in North Carolina and had served approximately two years in her current school and district. As it relates to the EVAAS, she had received a one-day training from another district in the previous

year. Regarding the NCEES, the participant did not disclose any training that had been provided. This participant was able to provide specific feedback about teachers whose scores were present in the district data, and this gave valuable input for this study.

Participant 6 was a White female who was in her fifth year as a principal in North Carolina. All five years were spent in her the district in which she worked at the time of the interview; all of these years were also at the same school. She stated that the district sent some administrators to EVAAS training in the past, but that she had not received any such training recently. At the time that she received the training, she felt that the professional learning was high quality. As it relates to the NCEES, she did not feel that she had received much training on this system. She had received some training from her district, but no outside training.

Participant 7 was an African American male who had been a principal for six years in North Carolina. At the time of the interview, he was in his fourth year in his current school district and in his third year at the school in which he was serving as principal. He felt that the quality of his training relating to the NCEES was somewhat lacking; he was not confident in his ability to translate the information to the entirety of his staff. He had to take it upon himself to seek out adequate information. He did not disclose information relating to his training on the EVAAS.

An overview of the demographic information of participants is contained in the table below. All principals were employed in middle schools in the Northeast or North Central regions of North Carolina:

**Table 4.1**  
*Participant Demographics*

	<b>Race</b>	<b>Gender</b>	<b>Years of Exp.</b>	<b>Years in Dist.</b>	<b>Years at School</b>	<b>F/R Lunch %</b>	<b>Rural/Suburban/Urban?</b>
Participant 1	White	Female	8	3	3	60.00%	rural
Participant 2	White	Female	7	7	1	70.00%	rural
Participant 3	White	Male	5	5	5	60.00%	rural
Participant 4	Black	Male	13	3	3	20.00%	rural
Participant 5	White	Female	6	2	2	100.00%	rural
Participant 6	White	Female	5	5	5	72.00%	rural
Participant 7	Black	Male	6	3	3	100.00%	urban
<b>Averages</b>			7.14	4	3.14	68.86%	

### **Research Questions**

The research questions for this study consisted of one quantitative question and four qualitative questions. The quantitative section of this study provides background data for the qualitative elements of the study. The data that were used for the quantitative section also provided context to the participants so that they were able to adequately answer the qualitative elements of the interview protocol.

#### **Quantitative Research Question**

To answer Research Question 1, I analyzed data that were provided by the North Carolina Department of Public Instruction (NCDPI). The responses to these research questions are succinct due to the nature of the data that were available. The charts and graphs that were provided by the NCDPI depicted the correlations between the ratings of middle grades mathematics teachers in participating districts in the North Carolina Educator Evaluation System (NCEES) and the scores of the same educators in the Educator Value-Added Assessment System

(EVAAS). These data depictions did not show exact numbers or percentages. For this reason, I had to make estimates based on my analysis of these data.

### ***Research Question 1***

Research Question 1 in this study was: *How do the correlations of NCEES and EVAAS ratings compare among the participating districts in the Northeast and North Central regions of North Carolina?* To respond to this research question, I reviewed the data regarding NCEES rating and EVAAS score correlations for middle grades math teachers that were sent to superintendents in response to my request to the NCDPI. These data were, in turn, sent to me by the superintendents in the participating districts.

NCEES Standards One and Four were chosen for the purposes of the data analyses that follow because these are the only two standards on which all teachers are rated regardless of their level of experience. Standard One is Teacher Leadership and gauges a teacher's ability to lead in their classroom, their school, and their district. Standard Four is Instructional Facilitation and addresses the ability of a teacher to adequately present academic content to their students. Standard Four was of particular interest in light of the fact that a teacher's ability to facilitate instruction should presumably yield a strong growth score as measured by EVAAS. While beginning teachers (teachers in years 1-3 of their careers) are rated on five different standards in NCEES, teachers who have more experience are only required to be rated on Standards One and Four.

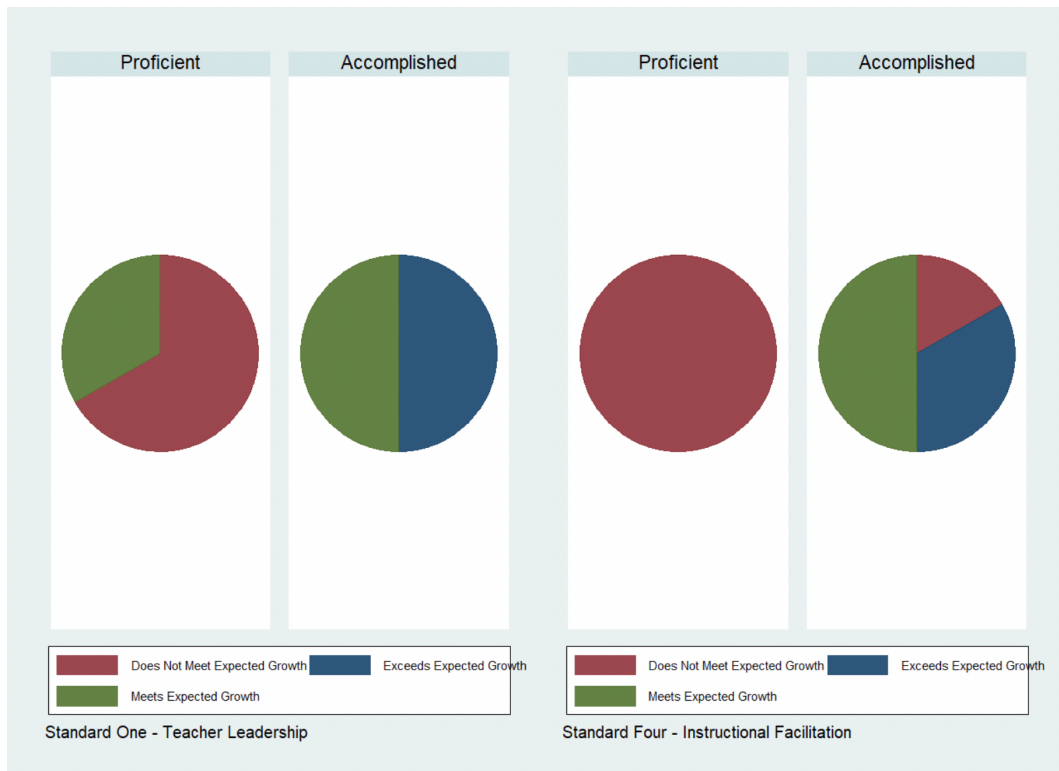
The EVAAS scores were expressed as numeric values that ranged from approximately -7 to +11. The graphs that were provided indicate how the ratings of teachers on the NCEES metric compared with their scores on the EVAAS scale. These graphs are detailed in the following content and are separated by district.

The pie graphs and box-and-whisker plots that are found in this section were provided to superintendents of participating school districts by the NCDPI, and were then provided to me by those superintendents. These graphs have not been altered and are in the original form they were when they were provided to me.

**District 1.** Figure 4.1 provides the NCEES rating and EVAAS score data correlations for District 1. Additional information that should be considered with respect to the two graphs that are discussed for District 1 is that the designations of Proficient and Accomplished are not the only ratings that exist in NCEES. In total, there are five categories for ratings of teachers in this system. Those five categories are: Not Demonstrated, Developing, Proficient, Accomplished, and Distinguished. What these graphs illustrate, therefore, is that none of the middle grades mathematics teachers in District 1 were assigned the lowest ratings of Not Demonstrated or Developing during that particular school year, nor were any assigned the highest rating of Distinguished.

**Figure 4.1**

*District 1 NCEES Rating and EVAAS Score Pie Chart Data*



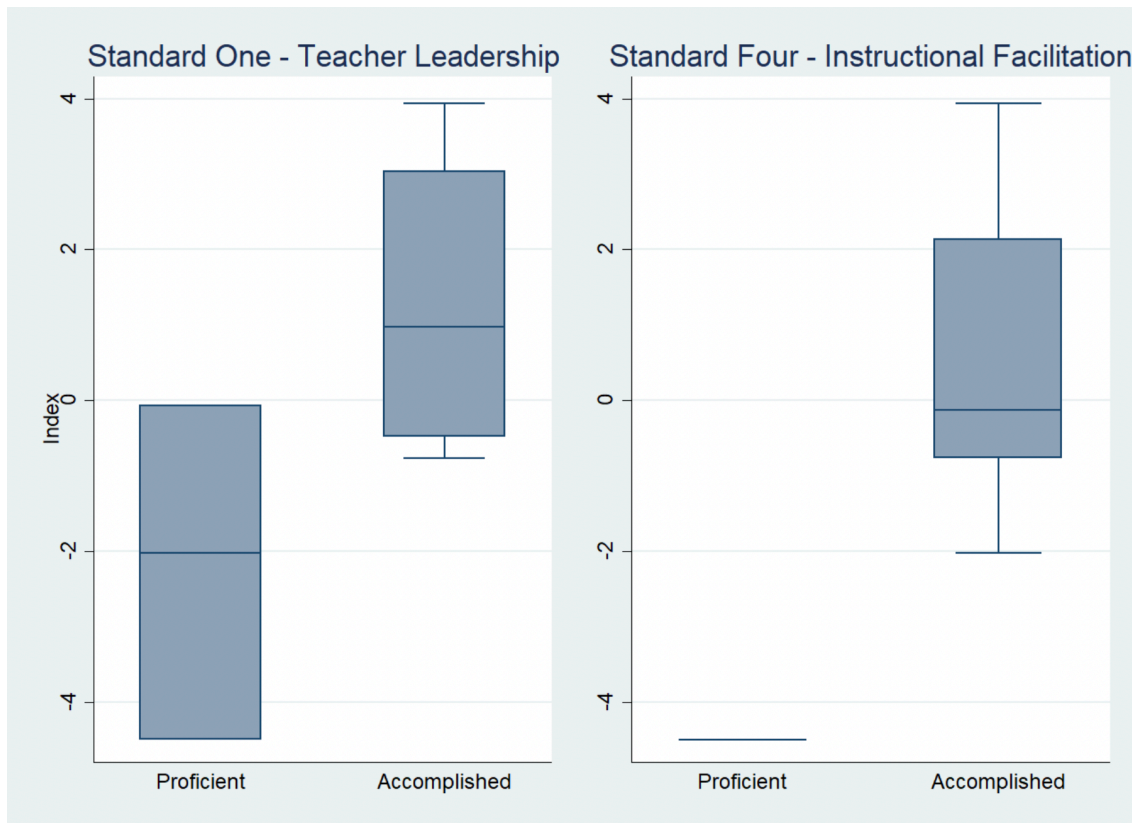
The images in Figure 4.1 show the percentages of middle school mathematics teachers who were rated Proficient or Accomplished in Standards One and Four and also scored Does Not Meet Expected Growth, Meets Expected Growth, or Exceeds Expected Growth on the EVAAS scale. For example, in NCEES Standard One – Teacher Leadership, approximately 67% of teachers who were rated as Proficient scored Does Not Meet Expected Growth on the EVAAS scale. For this same standard, approximately 33% of teachers who were rated as Proficient scored in the range of Meets Expected Growth on the EVAAS scale. The ratings of teachers who received the Proficient rating on this NCEES standard were, therefore, incongruent with EVAAS data that illustrate their actual impact on student achievement. The ratings of teachers who received the Accomplished rating on this NCEES standard were arguably congruent with EVAAS data that illustrate their actual impact on student achievement.

Figure 4.1 also provides the data for NCEES Standard Four – Instructional Leadership. In this instance, 100% of teacher who were rated Proficient in NCEES scored Does Not Meet Expected Growth on the EVAAS scale. The ratings of teachers who received the Proficient rating on this NCEES standard were, therefore, incongruent with EVAAS data that illustrate their actual impact on student achievement. For this same standard, approximately 20% of teachers who were rated Accomplished scored Does Not Meet Expected Growth in EVAAS, approximately 50% of teachers with the same rating in NCEES scored as Meets Expected Growth in EVAAS, and approximately 30% of teachers with that same rating scored as Exceeds Expected Growth in EVAAS. The ratings of teachers who received the rating of Accomplished on NCEES Standard Four were, therefore, largely congruent with EVAAS data that illustrate their actual impact on student achievement. The exception was teachers who received a score of Does Not Meet Expected Growth in EVAAS.

The second graph that was provided for each district is a box-and-whisker plot that indicates the ranges that these same teachers scored on the EVAAS scale as compared with their NCEES ratings. Figure 4.2 illustrates the range of scores that District 1 teachers received in EVAAS as compared to their NCEES ratings.

**Figure 4.2**

*District 1 NCEES Ratings and EVAAS Ranges Box-and-Whisker Data Plot*



This graph indicates that for teachers who were rated as Proficient in NCEES in Standard One – Teacher Leadership, the median rating was a -2 on the EVAAS scale. Their upper quartile score was also slightly negative, but approaching 0, and their lower quartile score was approaching -5 on the EVAAS scale. To provide some context regarding these scores, any score between -2 and +2 on the EVAAS scale would be considered as Meets Expected Growth. Any score above +2 would be scored as Exceeds Expected Growth, and any score below -2 would be scored as Does Not Meet Expected Growth.

Regarding Standard Four – Instructional Facilitation, this box-and-whisker plot shows that teacher(s) who were rated as Proficient on this standard in the NCEES, scored approximately -5 on the EVAAS scale, which would fall into the category of Does Not Meet

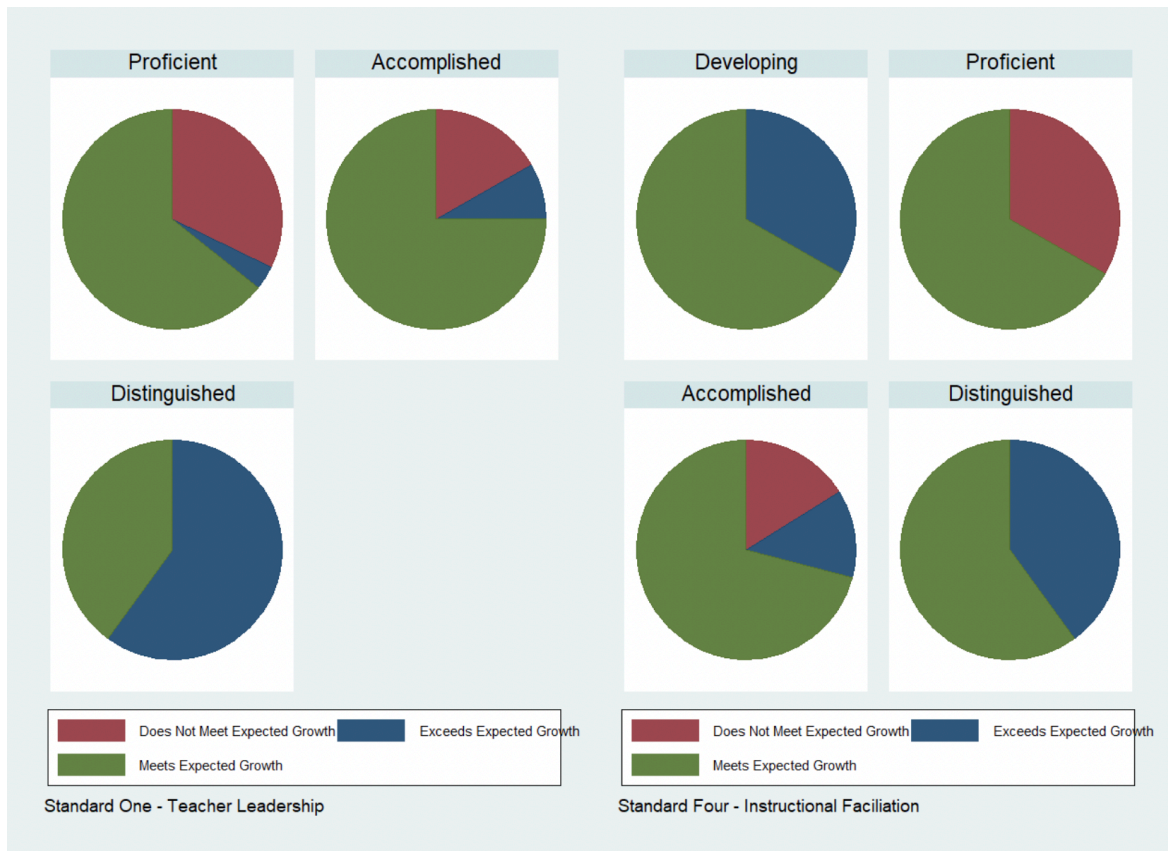


Expected Growth. It is important to point out that since there is only one line depicted for Standard Four on this graph, then the n-count for this data set was likely 1. This particular school district is quite small, so it is quite possible that only one teacher was rated as Proficient on Standard Four for that particular school year. The rating of the teacher(s) as Proficient in this standard appear incongruent with their EVAAs rating of Does Not Meet Expected Growth. One might assume that teachers who were rated as Proficient in NCEES Standard Four – Instructional Facilitation should at least have a score of Meets Expected Growth in EVAAS. For the teachers who were rated as Accomplished on Standard Four – Instructional Facilitation in NCEES, the median EVAAS score was slightly negative (between 0 and -1). The upper quartile was just under a score of +2, and the maximum score was just below +4. The lower quartile score for this group was approximately -1, and the minimum score was approximately -2. Again, any rating below -2 would be classified as Does Not Meet Expected Growth in EVAAS and anything above +2 would be classified as Exceeds Expected Growth in EVAAS. The data in this box-and-whisker plot indicate a loose correlation between the NCEES ratings and EVAAS scores of these teachers due to the subset of teachers who received a score of Does Not Meet Expected Growth yet were rated as Accomplished on NCEES Standard Four – Instructional Facilitation.

**District 2.** Figure 4.3 below shows the correlations between the NCEES ratings of middle grades mathematics teachers and the EVAAS scores of those same teachers for District 2. As with District 1, the graph depicts that there were no teachers who received an overall rating of Not Demonstrated in NCEES on either standard. The rating of Not Demonstrated is considered the lowest rating on the NCEES rubric.

**Figure 4.3**

*District 2 NCEES Rating and EVAAS Score Pie Chart Data*



For School District 2, in Standard One – Teacher Leadership, teachers were rated anywhere from Proficient to Distinguished. Of the teachers who were rated as Proficient on this standard, approximately 30% scored at the level of Does Not Meet Expected Growth on the EVAAS scale, approximately 67% scored Proficient on the EVAAS scale, and approximately 3% scored Distinguished on the EVAAS scale. For those teachers who were rated Accomplished on Standard One – Teacher Leadership, approximately 15% scored as Does Not Meet Expected Growth on the EVAAS scale, approximately 75% scored as Meets Expected Growth, and approximately 10% scored as Exceed Expected Growth. For teachers who were rated Distinguished on this standard, approximately 40% scored as Meets Expected Growth and

approximately 60% scored as Exceeds Expected Growth on the EVAAS scale. These graphics seemed to indicate that some correlation existed for these ratings of middle grades mathematics teachers, especially towards the upper end of the scale in Standard One – Teacher Leadership. The NCEES ratings for Standard Four – Instructional Facilitation also seemed to illustrate a similar correlation towards the upper end of the scale; however, it is important to point out that the teachers who were rated as Developing all scored either Meets Expected Growth or Exceeds Expected Growth in EVAAS. One might expect that teachers who were not considered Proficient in the area of Instructional Facilitation would likely have not done well on the EVAAS growth measures.

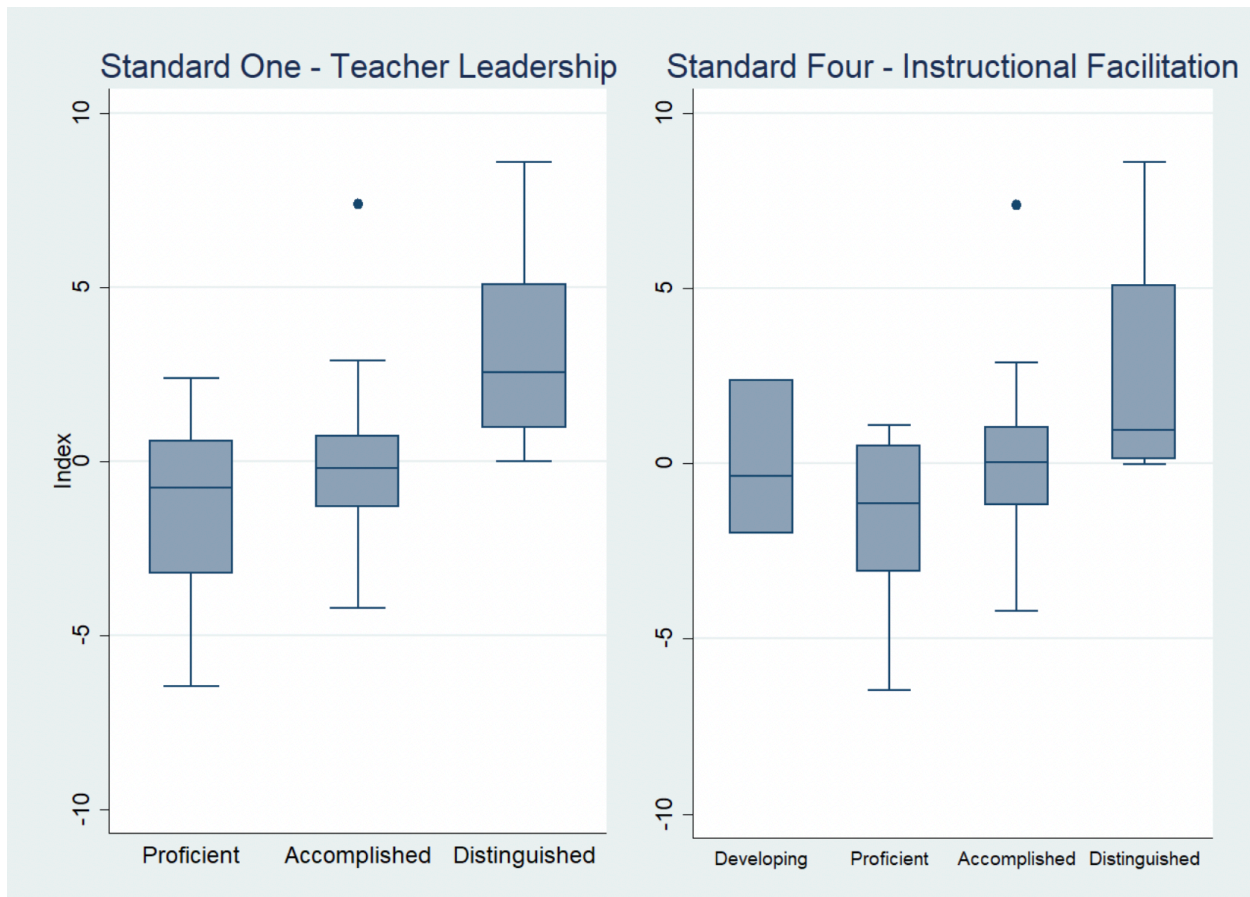
The right side of this graph shows the results for teacher ratings on Standard Four – Instructional Facilitation for this school district. In this standard, middle grades mathematics teachers scored anywhere from Developing to Distinguished for this particular school year. Of the teachers who were rated as Developing in Standard Four – Instructional Facilitation, approximately 60% scored as Meets Expected Growth on the EVAAS scale, and approximately 40% scored as Distinguished on the EVAAS scale. Among teachers who were rated as Proficient in Standard Four – Instructional Leadership, approximately 40% scored as Does Not Meet Expected Growth on the EVAAS scale, and 60% scored as Meets Expected Growth. For teachers who were rated as Accomplished on this standard, approximately 15% scored as Does Not Meet Expected Growth on the EVAAS scale, approximately 70% scored as Meets Expected Growth, and approximately 15% scored as Exceeds Expected Growth. For teachers who were rated as Distinguished on this standard, approximately 60% scored as Meets Expected Growth on the EVAAS scale and approximately 40% scored as Exceeds Expected Growth. The data depicted in these pie charts appear to show a correlation towards the upper end of the scale of both NCEES

and EVAAS. Teachers who were rated as Distinguished in NCEES also scored either Meets Expected Growth or Exceeds Expected Growth in EVAAS. However, there appears to be little correlation between NCEES ratings and EVAAS scores for teachers who were rated as either Developing, Proficient or Accomplished in NCEES. In particular, teachers who were rated as Developing on Standard Four – Instructional Facilitation also scored either Meets Expected Growth or Exceeds Expected Growth in EVAAS. One would assume that teachers who were rated as Developing on this standard in NCEES would have lower EVAAS scores.

As with District 1, the second graph that was provided for District 2 is a box-and-whisker plot that indicates the ranges that these same teachers scored on the EVAAS scale as compared with their NCEES ratings. Figure 4.4 illustrates the range of scores teachers received in EVAAS as compared to their NCEES ratings for District 2.

**Figure 4.4**

*District 2 NCEES Ratings and EVAAS Ranges Box-and-Whisker Data Plot*



The box-and-whisker plot above indicates how teachers' scores along the EVAAS scale compared to their scores in NCEES. The image on the left shows that teachers who were rated as Proficient in NCEES on Standard One – Teacher Leadership had a median score of approximately -1. Their lower quartile score was approximately -3, and their minimum score was approximately -7. Their upper quartile score was approximately +1, and their maximum score was approximately +3. For teachers who were rated Accomplished in NCEES on this standard, their median score was slightly negative (between 0 and -1). Their lower quartile score was approximately -2, and their minimum was approximately -4. The upper quartile score for this set of teachers was also approximately +1, and their maximum score was approximately +3, as well.

There was also one outlier for this group who scored approximately +8 on the EVAAS scale. Teachers who were rated Distinguished on this standard had a median score of approximately +3. Their lower quartile score was approximately +2, and their minimum was zero. Their upper quartile score was approximately +5, and their maximum was approximately +9. These data again illustrate a correlation of NCEES ratings and EVAAS scores towards the upper end of both measures, yet the ranges of EVAAS scores for teachers rated as Proficient compared to those who were rated as Accomplished are quite similar to one another.

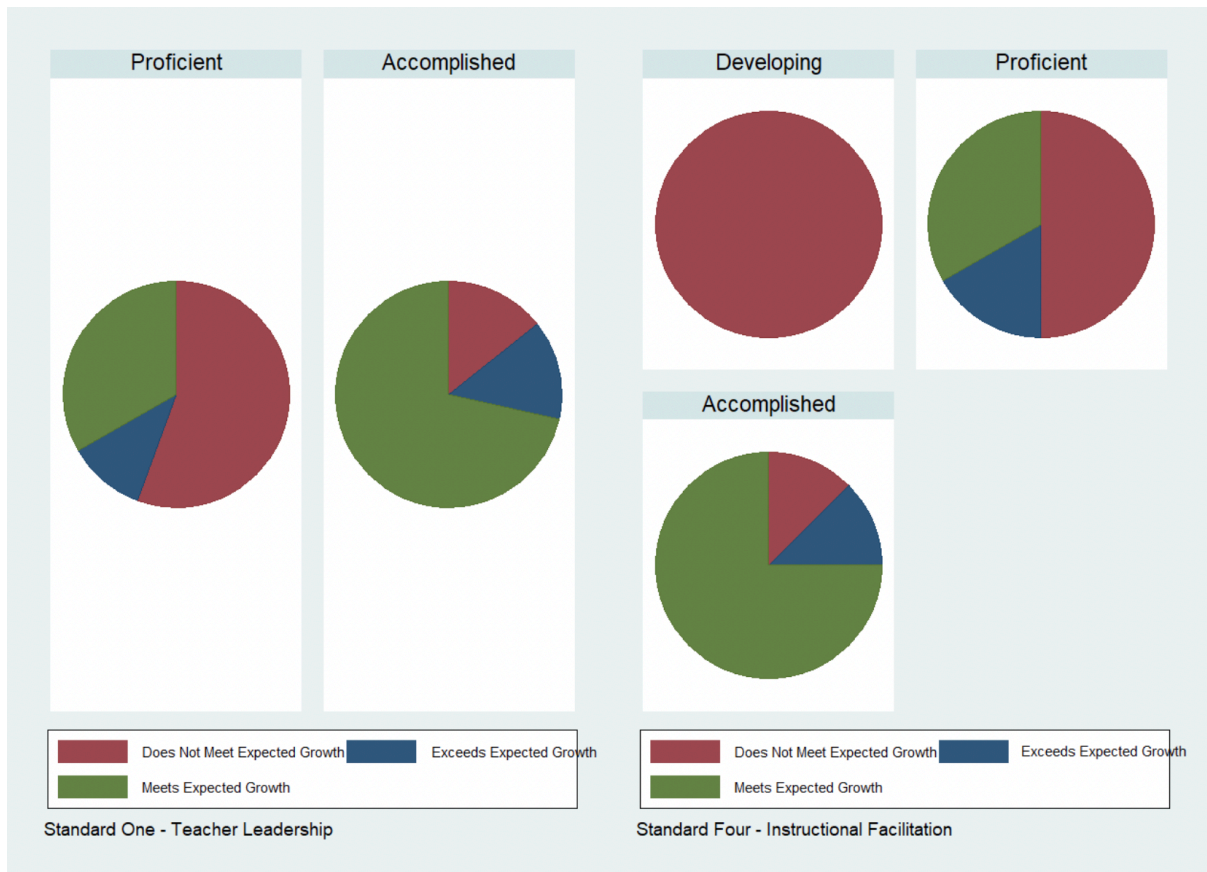
The graphic on the right shows that teacher who were rated Developing in Standard Four – Instructional Facilitation had a median score that was slightly negative (between 0 and -1). They had a lower quartile score of approximately -2, and an upper quartile score of approximately +3. For teachers who were rated Proficient on this standard, their median score was approximately -2. Their lower quartile score was approximately -3, and their minimum score was approximately -7. Their upper quartile score was approximately +1 and their maximum was approximately +2. Teachers who were rated as Accomplished in NCEES on this standard had a median score of approximately 0. Their lower quartile score was approximately -2 and their minimum was approximately -4. Their upper quartile score was approximately +2 and their maximum was approximately +3. There was also an outlier for this group who scored approximately +8 on the EVAAS scale. For teachers who were rated as Distinguished on Standard Four – Instructional Facilitation, their median score was approximately +2. Their lower quartile score was slightly positive (between 0 and +1), and their minimum score was approximately 0. Their upper quartile score was approximately +5, and their maximum score was approximately +9. These data also indicate a correlation between EVAAS and NCEES ratings of teachers toward the upper end of both measures, but the EVAAS scores of teachers

who were rated as Developing, Proficient or Accomplished seem to indicate incongruence. The teachers who were rated at the relatively low NCEES rating of Developing arguably performed as well or higher on the EVAAS scale than teachers who were rated as Proficient or Accomplished.

**District 3.** Figure 4.5 below shows the correlations between the NCEES ratings of middle grades mathematics teachers and the EVAAS scores of those same teachers for District 3. It can be seen here that District 3, as in the previous districts, did not have any teachers who had an overall rating of either Not Demonstrated or Distinguished in either of the standards. This means that middle grades mathematics teachers in District 3 did not receive either the lowest possible rating or the highest possible rating in NCEES.

**Figure 4.5**

*District 3 NCEES Rating and EVAAS Score Pie Chart Data*



The image above shows the data for the third participating school district. The pie graphs on the left side of the image indicate that, of the teachers who were rated as Proficient in Standard One – Teacher Leadership, approximately 55% scored as Does Not Meet Expected Growth on the EVAAS scale, approximately 33% scored as Meets Expected Growth on the EVAAS scale, and approximately 12% scored as Exceeds Expected Growth. Of the teachers who were rated as Accomplished in Standard One – Teacher Leadership in NCEES, approximately 15% scored as Does Not Meet Expected Growth in EVAAS, approximately 70% scored as Meets Expected Growth in EVAAS, and approximately 15% scored as Exceeds Expected Growth. From the data provided in these pie graphs, there appears to be very little correlation



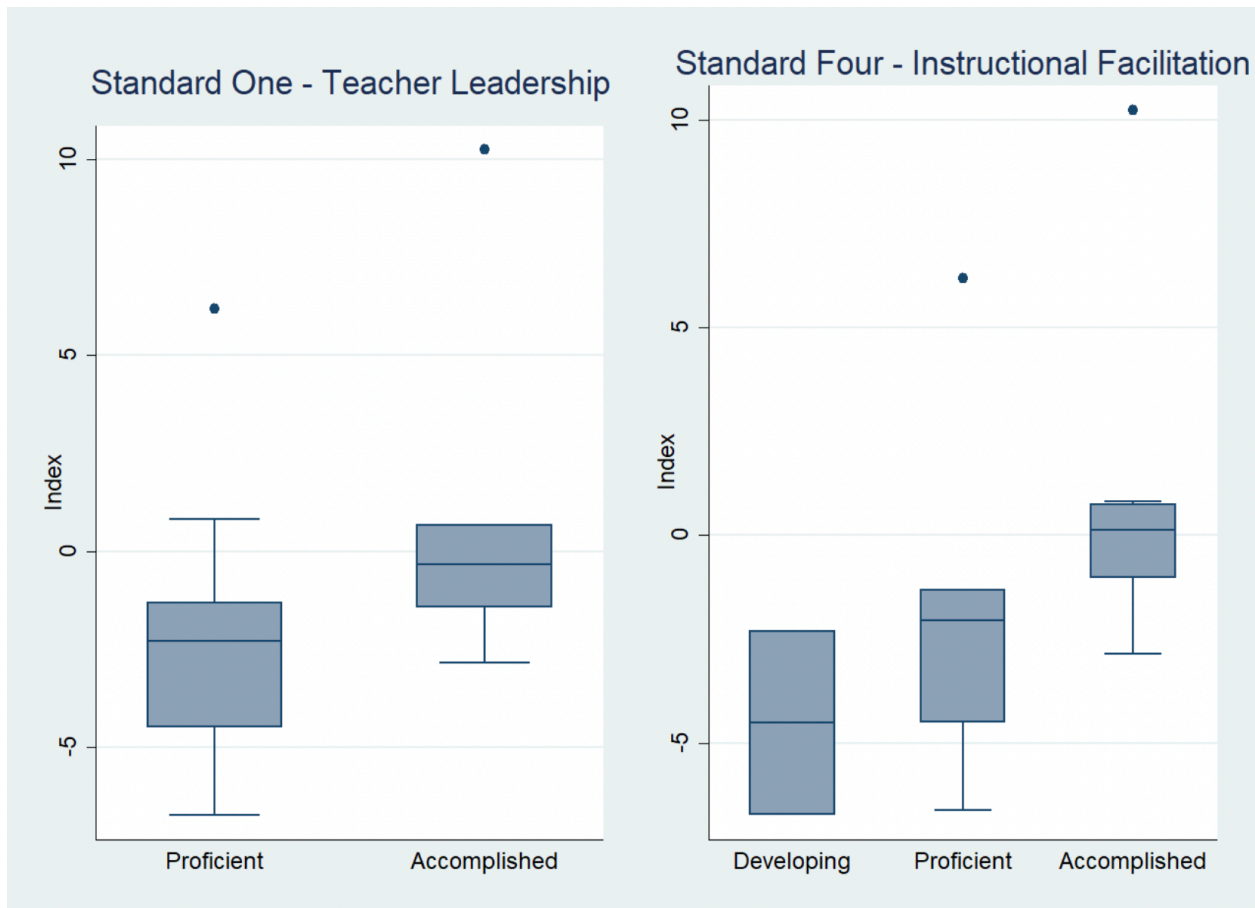
between teacher ratings in NCEES and their EVAAS scores for Standard One – Teacher Leadership.

The images on the right side of the graph show that, of the teachers who were rated as Developing in Standard Four – Instructional Facilitation, 100% scored as Does Not Meet Expected Growth in EVAAS. Of the teachers who were rated as Proficient on this standard, approximately 50% scored as Does Not Meet Expected Growth on the EVAAS scale, approximately 30% scored as Meets Expected Growth, and approximately 20% scored as Exceeds Expected Growth. Among teachers who were rated as Accomplished on Standard Four – Instructional Leadership, approximately 12% scored as Does Not Meet Expected Growth in EVAAS, approximately 75% scored as Meets Expected Growth, and approximately 13% scored as Exceeds Expected Growth. Again, it is important to note here that according to these graphics, there were no teachers who were rated as Not Demonstrated or Distinguished in NCEES for this district in either standard. At the lower end of both scales in Standard Four – Instructional Facilitation, there does appear to be some correlation between teachers who were rated as Developing in NCEES and scored Does Not Meet Expected Growth in EVAAS. One thing to keep in mind with these pie charts, however, is that since the data did not include the actual number of teachers who were rated a certain way in each standard, it cannot be determined if this apparent correlation exists because one teacher was rated in this manner, or several.

The second graph that was provided for District 3 is a box-and-whisker plot that indicates the ranges that these same teachers scored on the EVAAS scale as compared with their NCEES ratings. Figure 4.6 illustrates the range of scores teachers received in EVAAS as compared to their NCEES ratings for teachers in District 3.

**Figure 4.6**

*District 3 NCEES Ratings and EVAAS Ranges Box-and-Whisker Data Plot*



The box-and-whisker plots above indicate the range of scores in EVAAS as compared to their ratings in NCEES for middle grades mathematics teachers in this district. The image on the left shows that teachers who were rated as Proficient on Standard One – Teacher Leadership had a median score in EVAAS of approximately -3. This same group had a lower quartile score of approximately -4, and a minimum score of approximately -7. These teachers had an upper quartile score of approximately -2, and a maximum score of approximately positive +1. There was an outlier score for this group of approximately +6. For teachers who were rated as Accomplished in NCEES on this standard, there was a median score of approximately -1. The lower quartile score was approximately -2, and the minimum score was approximately -3. The

upper quartile score for this group was approximately +1. There was also an outlier score for this group of approximately +11. The range of scores shown in this box-and-whisker plot again indicates a lack of correlation with the NCEES and EVAAS ratings of this group of middle grades mathematics teachers in District 3. For all teachers in this group to have been rated either Proficient or Accomplished, yet their medians, lower quartile and upper quartile scores are all negative, or barely positive for the Accomplished group, seems to indicate a distinct lack of correlation. The two outlier scores appear to be the scores that make up the segment of the previous pie charts that indicated that some teachers exceeded expected growth for each of these groups. If only one teacher in each section scored as Exceeds Expected Growth, and the remainder of the group were scored mostly in the negative range, then these data give even greater evidence of a lack of correlation between the NCEES and EVAAS scores in this standard.

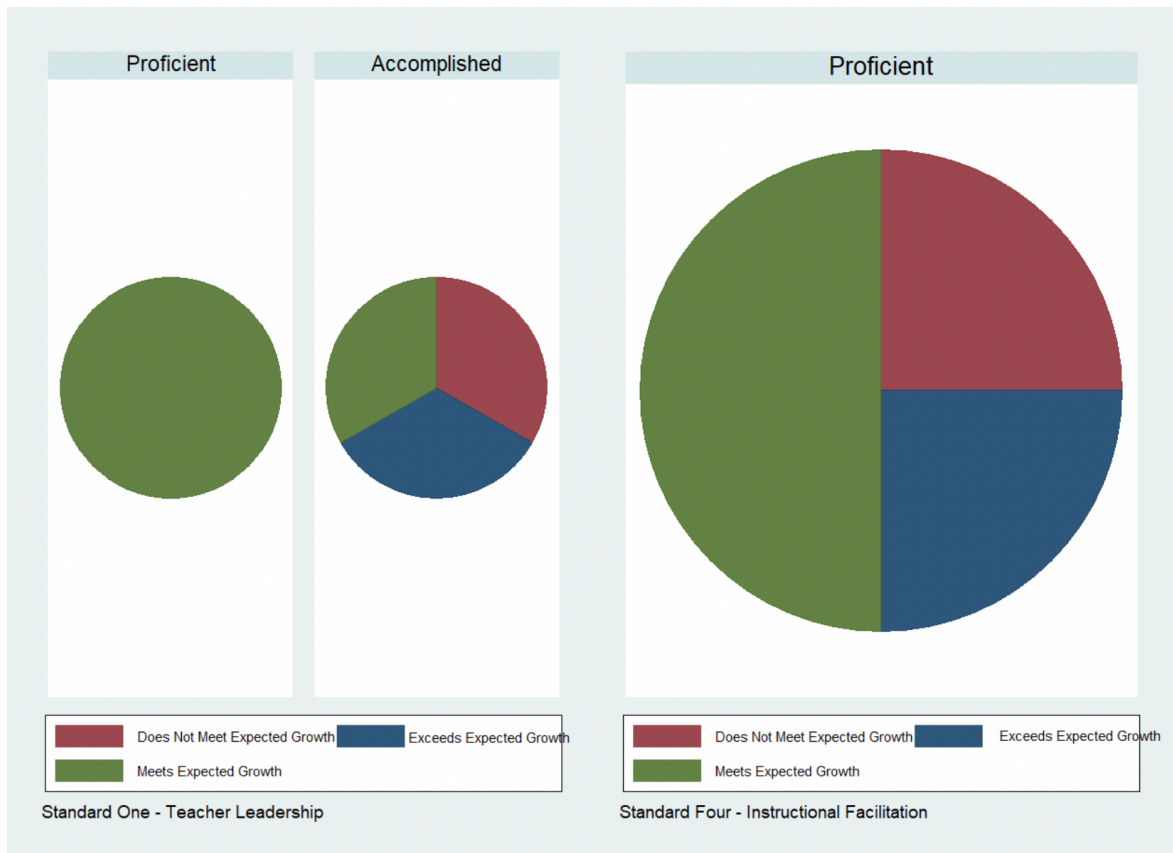
The right side of this graph shows that teachers who were rated as Developing in Standard Four – Instructional Facilitation had a median score of approximately -4 in EVAAS. The lower quartile score of this group was approximately -7, and the upper quartile score was approximately -3. Teachers who were rated as Proficient on this standard had a median score of approximately -3, a lower quartile score of approximately -4, and a minimum score of approximately -7. The upper quartile score for this group was approximately -2, and there was an outlier score for this group of approximately +6. Teachers who were rated as Accomplished on Standard Four – Instructional Facilitation had a median score of approximately 0. This same group had a lower quartile score of approximately -1, and a minimum score of approximately -3. The upper quartile score for this group was approximately +1, and the maximum score was approximately slightly higher than +1. There was also an outlier score for this group of

approximately +11. While there does appear to be some correlation between teachers who were rated Developing on this standard and also receiving an EVAAS score of Does Not Meet Expected Growth, the box-and-whisker plot for teachers who were rated as Proficient in Standard Four – Instructional Facilitation appears to be very similar to the box-and-whisker plot for teachers who were rated as Developing in this standard.

**District 4.** Figure 4.7 below shows the correlations between the NCEES ratings of middle grades mathematics teachers and the EVAAS scores of those same teachers for District 4. It can be seen here, as in the previous districts, that a very limited number of the available ratings for teachers in NCEES were actually used by this district during the school year for which the data were provided. There were no teachers who were rated at the Not Demonstrated, Developing, or Distinguished performance levels on either standard. For Standard Four – Instructional Facilitation teachers were all rated as Proficient.

**Figure 4.7**

*District 4 NCEES Rating and EVAAS Score Pie Chart Data*



In the pie graphs for District 4, the NCEES scores of middle grades mathematics teachers are compared with their EVAAS scores. For teachers who were rated as Proficient in Standard One – Teacher Leadership, 100% scored as Meets Expected Growth on the EVAAS scale. Of the teachers who were rated as Accomplished in this standard approximately 33% scored as Does Not Meet Expected Growth, approximately 33% scored as Meets Expected Growth, and approximately 33% scored as Exceeds Expected Growth on the EVAAS scale. There does appear to be some correlation between the proportion of teachers who were rated as Proficient in Standard One – Teacher Leadership and an EVAAS score of Meets Expected Growth. The teachers who were rated as Accomplished, however, appear to have scored evenly among all

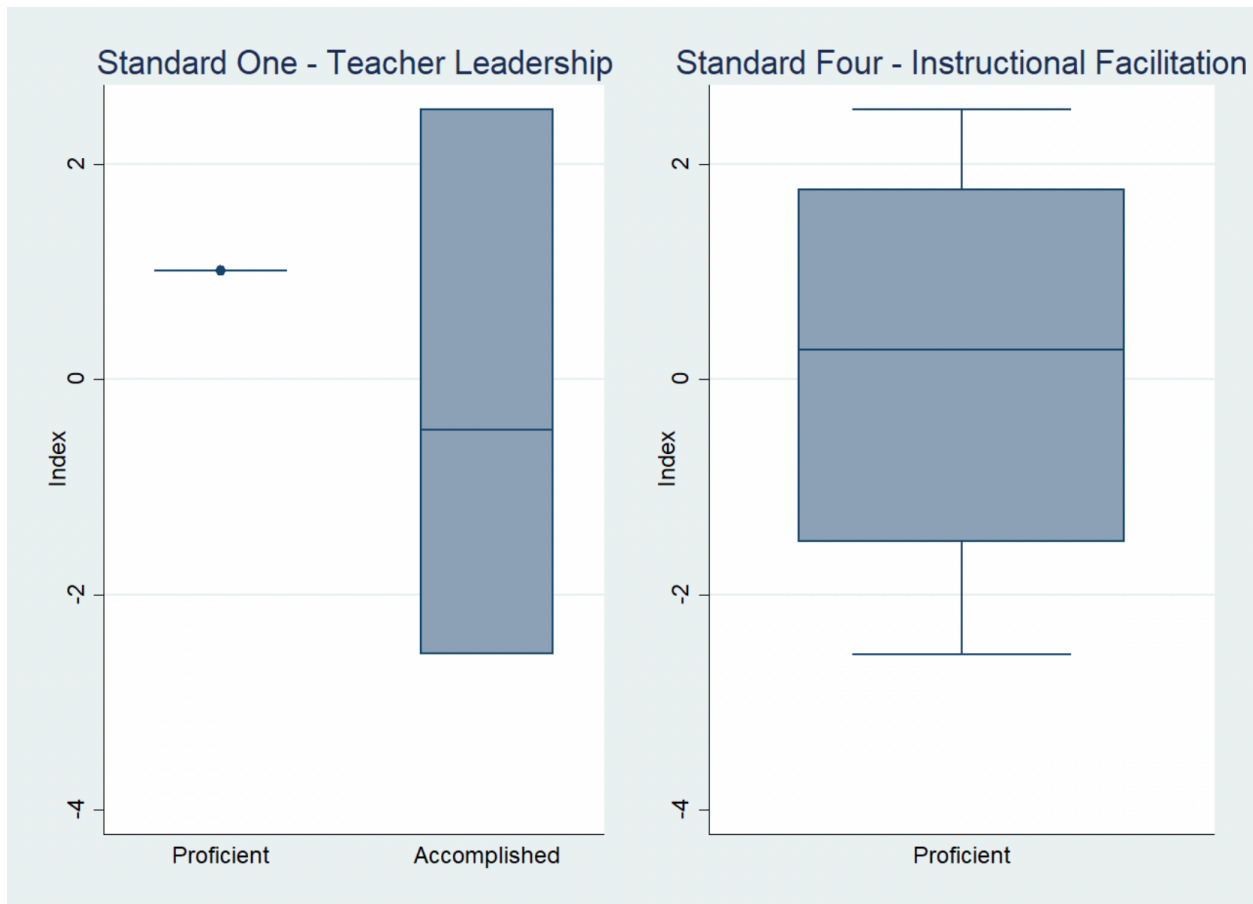
three categories that are available in EVAAS. This would seem to indicate that there is very limited correlation between the proportion of teachers in this district who were rated as Accomplished during the school year and their respective EVAAS scores.

The right side of the graphic in Figure 4.7 shows that of the teachers who were rated as Proficient on Standard Four – Instructional Leadership, approximately 25% scored as Does Not Meet Expected Growth, approximately 50% scored as Meets Expected Growth, and approximately 25% scored as Exceeds Expected Growth. From the information in this graphic, one can see that the only ratings that were used in NCEES for these standards were Proficient and Accomplished. This district is a very small school district, so it is important to note that the n- count for these data is likely very small. Based on the data shown in Figure 4.7, there would appear to be little to no correlation between how teachers were rated on Standard Four – Instructional Facilitation and their respective EVAAS scores. Since all teachers were rated the same in NCEES, but their EVAAS scores were approximately 25% Does Not Meet Expected Growth, 50% Meets Expected Growth, and 25% Exceeds Expected Growth, it would appear that the NCEES ratings in District 4 do not predict how middle grades mathematics teachers will score on the EVAAS scale.

The second graph that was provided for District 4 is a box-and-whisker plot that indicates the ranges that these same teachers scored on the EVAAS scale as compared with their NCEES ratings. Figure 4.8 illustrates the range of scores teachers received in EVAAS as compared to their NCEES ratings for District 3.

**Figure 4.8**

*District 4 NCEES Ratings and EVAAS Ranges Box-and-Whisker Data Plot*



The box-and-whisker plots above show the range of EVAAS scores teachers received in this district compared with the NCEES scores of the same cohort group. On the left side of the graphic, it shows that teachers who were rated as Proficient in Standard One – Teachers Leadership had a median score of approximately +1. There are no lower or upper quartile scores indicated for this group, nor any minimums or maximums. Teachers who were rated Accomplished on this standard had a median score of approximately -0.5, a lower quartile score of approximately negative -2.5, and an upper quartile score of approximately positive +2.5. Figure 4.8 does appear to indicate a correlation between teachers who were rated as Proficient and their respective EVAAS scores. However, with the lack of variability that is present on this

box-and-whisker plot, the number of teachers who were rated in such a way is likely very small. It is quite possible that the data shown in this section is only representative of one teacher. For teachers who were rated as Accomplished on this standard, these data would suggest that there is no correlation between this rating and the respective EVAAS scores of teachers.

The image on the right side shows that, of the teachers who were rated as Proficient in Standard Four – Instructional Leadership, the median score was approximately +0.5, the lower quartile score was approximately -1.5, and the minimum score was approximately -2.5. The upper quartile score for these teachers was approximately +1.5, and the maximum score was approximately +2.5. Since all middle grades mathematics teachers were rated the same in NCEES for District 4 in the school year for which the data were provided, yet their EVAAS scores ranged across all possible ratings in EVAAS, I expected there to be a great lack of correlation displayed on this box-and-whisker plot. When I looked a little closer, however, I noticed that the teachers who were rated as Does Not Meet Expected Growth just barely fell into that rating category. The same is true for teachers who received an EVAAS score of Exceeds Expected Growth. As described earlier in this section, for teachers to receive a score of Does Not Meet Expected Growth, their score must fall below -2 on the EVAAS scale. In the same manner, for teachers to receive a score of Exceeds Expected Growth, their score must be above +2. The minimum and maximum scores in the box-and-whisker plot above for teachers rated as Proficient in Standard Four – Instructional Facilitation fell just below -2 and just above +2. The remainder of the data points appear to have fallen within the range of -2 and +2, which would indicate a smaller degree of variance than I initially anticipated.

In summary, the findings for Research Question 1 largely suggest that some level of correlation may exist among the upper levels of both rating systems. To a lesser degree some



correlation exists on the lower extreme; however, the lowest rating in NCEES, Not Demonstrated, did not appear in any of the graphics from participating districts. This would indicate that no teachers received Not Demonstrated as an overall rating in NCEES from any of these districts.

Teachers who were rated as Distinguished in NCEES tended to outperform their counterparts in their EVAAS growth scores. However, for the lower and middle ranges of ratings in NCEES, there appeared to be very little correlation with the EVAAS scores of teachers who were rated in such a manner. It is important to point out that participants were only provided with the data for their own districts when responding to the questions in the interview protocol.

### **Qualitative Research Questions**

To answer Research Questions 2-5, I conducted interviews with seven participants who served as principals in the Northeast or North Central Regions of North Carolina. After completing the interviews, I reviewed and analyzed participant responses to the interview questions and identified emergent themes that enabled me to address each research question. Due to the limited number of participants in this study, a threshold was not utilized to quantify a theme.

To provide background to participants in preparation for their interview questions, I provided two data sets for their review. These data sets were specific to each participant's district and outlined a) the ratings that middle grades mathematics teachers received from administrators on Standards 1 and 4 in NCEES and b) the growth ratings that the teachers received from EVAAS. The themes that emerged from their responses are described in the following paragraphs.

## ***Research Question 2***

Research Question 2 in this study was: *How do principals describe the levels of correlation that exist between their district's NCEES and EVAAS ratings?* As I reviewed the participant responses from the interviews that were conducted, themes emerged that helped me to answer this research question.

### **Expectation of Some Correlation between Standard 4 in NCEES and EVAAS**

**Ratings.** One theme that emerged from the participant responses was that it is feasible that the ratings of teachers on Standard 1 in NCEES would not align with their EVAAS ratings. Standard One in NCEES is titled “Teachers Demonstrate Leadership.” Based on the responses from participants, teachers can demonstrate leadership in a variety of ways that may not necessarily yield results in the area of student growth in academics. From the participants’ perspectives, Standard One in NCEES and teacher results in EVAAS may be measuring different things, thus there may not be very much correlation between the two. Regarding Standard One, one participant pointed out:

...the actual document that we use for making observations, doesn't align. So I can come in under standard one and I can do all of this. Start on time, I'm using the standards, I'm using formative assessments and I'm really leading in my classroom and I know my data and yet at the end of it, I might not meet growth... I may go in and see that someone's got kids in groups and yet the groups might not be meaningful in that the student dialogue might not lead to a deeper understanding. So therefore I've ranked this person as either proficient or accomplished, but the depth of knowledge that the kids came out with doesn't impact EVAAS.

The absence of correlation, therefore, should not necessarily be surprising.

On the other hand, Standard 4 in NCEES is titled “Teachers Facilitate Learning for their Students.” For this standard, participants did expect there to be a greater degree of correlation between the teacher ratings in NCEES and the teacher ratings in EVAAS. During our interview, one participant did point out that the questions in Standard 4 of NCEES do not “get into high yield instructional practices.” One participant also pointed out:

I don't believe that leadership, the standard of leadership necessarily translates into the facilitation of learning per Standard Four, I do think that Standard Four and EVAAS should be more closely aligned.

**Little to No Correlation between NCEES and EVAAS.** Regardless of the district in which each participant served, they made consistent observations about the lack of correlation that existed between NCEES and EVAAS from the data that was provided by the NCDPI. They also provided some insight in their interviews as to why that could be the case. Many of the participants reported that, as principals, they had received very little training on either system of performance metrics. Some participants expressed that they had participated in professional development for either NCEES or EVAAS early in their administrative tenure, but not both. From that point forward, they added, little to no training on these systems had been provided. Participants also stated that, while they had not received much training on the actual systems of evaluation, they were involved with conversations at the district level where their district leaders had pointed out that their ratings of teachers in NCEES should align with the respective teacher’s EVAAS ratings. However, little guidance was given other than this general expectation. One participant did state that her district went through a process of reviewing each standard rating in NCEES and discussed with their colleagues the evidence they should see in the classroom in order for a teacher to be rated at each level (Not Demonstrated, Developing, Proficient,

Accomplished, Distinguished). That training took place when NCDPI first transitioned to using NCEES and had not been conducted since. One participant observed:

There's been no professional development in terms of where's the alignment, where are the pieces of NCEES that directly impact student achievement, and how can those things be matched up? There's been no conversations about that. It's just, 'Here's the data.' We all think that they should probably align better and so then it's kind of understood that the principal should go back and maybe pay more attention that when they're completing evaluations. Keep that in the back of your mind, is this teacher someone who meets or exceeds growth on a regular basis, or is this a teacher that does not meet growth? And keep that in the back of your mind as you're filling out your evaluations.

While the majority of participants did not perceive that the NCEES and EVAAS ratings of middle grades mathematics teachers in their districts were correlated, a couple of participants differed, at least in part, based on the district level data we reviewed. One participant believed that there was some correlation between the higher range of the ratings for his district; these ranges include the rating of “Accomplished” in NCEES and the ratings of “Meets” or “Exceeds Growth” in EVAAS, specifically stating:

on the upper end, on the distinguished end, there was more correlation, still not a perfect correlation...I would give that somewhere on, if we're talking about correlation coefficients...I mean, that's a medium strength correlation, I would say I saw there. In terms of like the proficient and developed on the lower end, I mean, there's no correlation...the results to me look kind of all over the place.

There was another participant response that was an outlier that is worthy of note; this participant responded that the data sets were well-aligned in his district.

I think they align perfectly. In trainings, we were always told that if your teacher meets or exceeds, it kind of goes along with your accomplished, distinguished, your ratings and EVAAS...And then if they're not meeting, it should align with their being developing or proficient, somewhere like that. And so they were kind of giving us examples of what it looks like, and when you really look at it, it was true.

To summarize, many participants believed that there should be some level of correlation between teacher ratings in NCEES and their EVAAS scores. However, little to no correlation existed with the data that were provided for the majority of the districts. However, some correlation seemed to be apparent in certain districts at the very low or very high ends of the rating scales.

### ***Research Question 3***

Research Question 3 in this study was: *How do principals describe the levels of correlation that exist between their district's EVAAS and NCEES ratings compared to their daily observations of teachers within their school?* As I reviewed the responses of participants, themes emerged that helped me to answer this question. Those themes are discussed in the following paragraphs.

**Correlation at the Extremes.** When looking at the groups of teachers who were rated as Developing on either standard, there appeared to be some alignment with those teachers being rated as either Does Not Meet Growth in EVAAS or Meets Expected Growth. For teachers who were rated as Distinguished, there was some correlation with EVAAS ratings of either Meets Expected Growth or Exceeds Expected Growth. Participants pointed out that for teachers to receive an overall rating of Developing on a standard in NCEES, then their performance is likely, as one principal observed, “really, really bad.” Another participant elaborated on this a bit

further and stated that if a teacher receives an overall rating of Developing in a standard, then that teacher has to be moved to an improvement plan. Such a circumstance obligates the principal to offer strategies and support in hopes that the teacher might improve the Developing rating to a better performance rating. One participant's perspective about this issue seemed to be something up: "...there's all kinds of things that you have to do..." when working with a teacher through this process.

With regard to the correlations that were seen at the upper level ratings, one participant pointed out that any such correlations could have been attributed to conversations between principals and district personnel. In such conversations, this principal suggested, they were probably encouraged to ensure that NCEES and EVAAS ratings aligned. During these conversations, however, it was also likely that no guidance was provided in terms of what to actually look for within the NCEES standards. She specifically stated, "...I certainly would not rate any correlation to any sort of professional development or training that we've received from the district, because ...none of that has happened..."

**NCEES Can be Easily Manipulated.** Many participants mentioned that principals have a lot of flexibility and autonomy when it comes to teacher ratings in NCEES. As previously mentioned, if teachers have an overall rating of Developing in a standard, then it is expected in many districts that those teachers be moved to a performance improvement plan. If a principal has to do that, then there are additional expectations placed on the principal for the upcoming school year. One participant stated:

...it's always pretty [much] a hundred percent that if you marked them developing, they're going to be, does not make growth [in EVAAS]. And I think that's just because we don't mark teachers developing for the year, unless it's a big problem, because then

you [have] to move on to a plan...if you're not able to coach somebody out of a directed [plan]...it's pretty serious. You know, even a beginning teacher...you have to move them. One participant even stated rather bluntly that "...if there's a teacher [who] shouldn't be [rated] very high, you can ensure they don't rate very high."

Based on the responses of the participants, it can be summarized that teacher ratings in NCEES are, as one participant opined, very "open to interpretation" by the evaluator. When such ratings are compared with EVAAS scores that are strictly based on assessment data, one could see why there may be a disconnect between the correlations of these two evaluation systems.

#### ***Research Question 4***

Research Question 4 in this study was: *To what factors do principals in the Northeast and North Central regions of North Carolina attribute the level of correlation between the NCEES and EVAAS ratings of teachers?* As I reviewed participant responses to the interview protocol, themes emerged that provided insight to Research Question 4. These themes are discussed in the following paragraphs.

#### **School and/or District Level Interventions Not Factored into Teacher Ratings.**

Participants pointed out that there are various things happening in a school that could impact student achievement that are not necessarily directly related to the teacher, and thus would not show up in their NCEES ratings. One participant pointed out that during the school year for which the data on teacher evaluations were provided, their district had adopted a new mathematics curriculum. Teachers were following the new curriculum, but this participant was not confident that the pacing of the curriculum allowed for all standards to be covered that would have been tested on the standardized test at the end of the year. He explained:

...it would have been the first year of implementation for that math curriculum, so I do think that could impact it also. Because I think if the teachers followed the program and were trying, their evaluations may be rated much higher than the actual outcome for the students because they were hitting some of the strategies and they were doing well, but they weren't familiar enough with the pacing of it...the appropriate coverage of the standards and the correlation to the test specifications and what was tested and how it was tested, we were still figuring a lot of it out then.

Another school level factor that was mentioned by participants is the Multi-Tiered System of Support (MTSS) that exists in most schools. The MTSS model is one in which a teacher can come to a group of their peers and ask for feedback and suggestions on how to help a particular student. That teacher is then provided with information and materials that have proven to be beneficial for their colleagues. These additional strategies may have impacted student achievement but may not have necessarily been present during the day the teacher was evaluated in the NCEES system. Sometimes this model includes having other staff members work with a particular student, or group of students, to ensure that students receive as much support as possible. This also would not be factored into a teacher's NCEES rating. When discussing the MTSS process, one participant pointed out that principals may hire "interventionists" to help students who are struggling. The way this participant described interventionists would align with what many would consider a tutor. The interventionists work with students who have struggled academically and need additional support. This support is typically provided during the school day. She described how both the MTSS model and the interventionists could impact NCEES and EVAAS correlations in the following way:



Some of the growth the teacher gets credited for is the teacher, but at the same time, every little thing is not accounted for. We do account for EC, we do account for interventions. We try to do the, when you do the roster verification, do the percentage of time they're with somebody, but then it kind of shares that credit where maybe the EC teacher was super effective or the interventionist was super effective. So that's the time I don't always see them correlating.

Simply put by another participant, those “attributes are not measured in this document, in my professional opinion.”

**NCEES is a Very Subjective Rating Scale.** Participants consistently indicated that different evaluators could enter a teacher’s classroom and that the NCEES ratings of the teacher’s performance could vary in each standard from evaluator to evaluator. These responses indicate that NCEES ratings are, in the view of these principals, very subjective. The language that is used in the rubric in NCEES was described as more “generalized” by one participant, who further asserted that this could lead to such subjectivity.

I could go in and I could say this person's proficient, not accomplished because I tend to think of proficient [as] doing your job. I'm giving the kids a year's worth of growth, I'm aligning the core one so that it's strong enough that, that kid can meet or exceed growth. Another administrator may go in because of the subjectivity in the actual platform and they may say, well that person's accomplished because they did the group. Well, the specificity in the platform might say, well the kids are managing and leading their own group. That's open to interpretation. So it's a very general subjective document to me, not aligned with EVAAS.

The participants indicated that some level of subjectivity also could be attributed to the experience level of the teacher. This would not necessarily be the case because of how the teacher performs, but rather, might arise from the mindset of the administrator who is evaluating teachers. These evaluators sometimes have different expectations of what they would like to see in a classroom from a beginning teacher and a veteran teacher. One participant specifically stated:

...I think another way that factors for our district when you look at that is we do have a lot of beginning teachers...I think that's where the subjectivity comes in, because they don't have to be proficient until X number of years, but it's hard to have the same lens going into a BT's classroom that you have going into a teacher with 10 or 12, 15 years of experience. So sometimes I think the scores could be inflated for even some of our BTs because it becomes so personal, the teacher evaluation process does.

From the responses of these participants, it seemed that it may not only be the evaluation instrument creating the subjective nature of the ratings, but also the mindset of principals and other administrators who are conducting these evaluations.

To summarize the findings from Research Question 4, the correlation, or lack thereof, may be impacted by interventions that are put in place by either a school and/or district, and by the subjective nature of the evaluation process. With various factors not being accounted for in NCEES, most participants in this study were not necessarily surprised by the lack of overall correlation between NCEES and EVAAS ratings.

### ***Research Question 5***

Research Question 5 in this study was: *What teacher attributes do principals in the Northeast and North Central regions of North Carolina believe would increase student*

*achievement and thus produce a stronger correlation in NCEES and EVAAS ratings, if an evaluation of these attributes existed in NCEES?* As I reviewed participant responses to the interview protocol, themes emerged that provided insight to Research Question 5. Those themes are discussed in the following paragraphs.

**Relationships.** During the interview process, participants consistently mentioned the relationships that teachers are able to build with students as one of the attributes that they believe impact student achievement. They acknowledged that the exact nature of good relationships with students can be “hard for people to be able to put into words,” it is something that most participants stated is an important factor that contributes to an increase in student achievement. One participant even believed that strong relationships could overcome many other important factors that would traditionally be looked for in a teacher.

...I think one of [the factors] is relationships with students, [and] their ability to grow relationships with students. I've got one teacher who builds great relationships with students. Isn't probably the best deliverer of all the information. Doesn't always follow the pacing guide. Is terrible about coming to work and doing grades and some of those other things, but the guy meets growth every year just because I think he has really good relationships.

One participant felt so strongly about teacher-student relationships and their impact on student achievement that, based upon observations during his career, they were “directly correlated” with EVAAS. It is also important to point out that teacher-student relationships were the first factor mentioned by many participants when asked about teacher attributes that contributed to student achievement. Multiple participants described such relationships as “first and foremost” or “number one” when considering teacher attributes that are linked to student achievement.

**Pedagogy.** The other theme that emerged from the responses of participants in this study revolved around pedagogy; their definition of this term focused largely on a teacher's ability to present content to students in a way that students understand. There were a few different sub-elements of pedagogy that were discussed by participants, and those elements are detailed in the following paragraphs.

One sub-element of pedagogy that was discussed was whether or not teachers had the ability to align their instruction to the North Carolina Standard Course of Study (NCSCOS). A teacher's ability to understand the standards, they observed, may not necessarily translate to that teacher being able to disseminate the information in a way that enables students to grasp the concept. One participant put it this way:

[NCEES] talks about the standards, but it doesn't really talk about aligning the standards with the lesson...it basically is like, are they planning? Are they looking at standards?...but just because you're looking at standards, or you have standards listed, doesn't mean that's really what they're assessing or really what they're facilitating...I think the alignment piece of instruction could be better addressed in NCEES.

Another sub-element of pedagogy that was mentioned by participants was how a teacher questions in the classroom; this observation included the types of questions and whether or not they are making sure that all students have an opportunity to respond to questions in the classroom. One participant provided a description of what she looks for specifically with questioning and described it this way:

...the questioning of the teacher, level of question, number of questions, amount of questions, how they ask the questions, using equitable strategies like cold calling and

equity sticks and things like that, versus raised hand questions so that everyone's participating.

A third element of pedagogy that was discussed by participants was whether or not students were actively engaged in the lesson being taught. Participants described student engagement in a variety of ways. Some of those ways were: “facilitation of learning,” “participation in conversations or discussions,” “teach[ing] the students in a manner which they can understand,” and “math talk.”

To summarize the findings from Research Question 5, three main themes emerged regarding teacher attributes that participants felt contributed directly to student achievement. Those three themes were the relationships that teachers had with students, the ability of the teacher to question students appropriately, and the pedagogical knowledge that teachers possess. Participants believed that if these attributes were addressed more directly in NCEES, the correlations between NCEES and EVAAS ratings would be much higher. There was a notable exception; One participant stated that there would still likely be a discrepancy because NCEES would still be subjective. This individual stated specifically, “It won't be a difference because basically, in NCEES...[it's] still going to come down to someone else's observation....” An additional participant felt that there should be some correlation if these attributes were addressed more specifically but noted that until there is consistent training from upper level leadership on what to look for from these standards, there would still be some misalignment.

I think it depends...theoretically it should [strengthen the correlation], but until there's consistent training and expectation and monitoring of the NCEES system, it's not going to...If a principal...evaluate[s] differently than I evaluate, and you understand...what you think makes a good teacher, or engage the students might be different than what I think.

And until we sit down and have a conversation about it, we can see the same person and not rate the same. And it's, that it's not going to correlate or very loosely correlate with EVAAS.

### **Chapter Summary**

This chapter presented the findings of the study that was conducted using data provided by the NCDPI and data from the interviews. The findings of the study were organized around the research questions and the themes that emerged from the participants' responses. The questions to which these participants responded included background questions about their experience as an administrator, as well as their perspective on EVAAS and NCEES correlations. They also discussed the degree of training they had received on both evaluation measures along with the teacher attributes that may contribute to an increase in student achievement that may not already be listed in the NCEES rubric.

## **CHAPTER 5: DISCUSSION**

### **Chapter Introduction**

Chapter 5 provides a discussion of the findings from Chapter 4 and the conclusions drawn from the study. In this chapter, I also provide implications of the results of this study for policymakers and practitioners. The chapter concludes with suggestions for future research.

### **Purpose of the Study**

For this study, I used a qualitative design to gain insight into the perspective of principals from the Northeast and North Central region of North Carolina about the relationship between teacher ratings on the North Carolina Educator Evaluation System (NCEES) and their scores on the Educator Value-Added Assessment System (EVAAS). I conducted virtual interviews with principals from these regions; I searched for emergent themes once those interviews had been transcribed. The study examined principals' opinions of their level of trainings with regard to NCEES and EVAAS evaluations, and whether they believed those evaluation measures were aligned with one another.

The evaluations that are conducted in the NCEES are considered to be subjective in nature. Principals rate teachers using a rubric that is provided by the North Carolina Department of Public Instruction (NCDPI). With this tool, teachers can be rated using the hierarchical categories of performance; Distinguished serves as the highest rating level.

EVAAS is used to measure student growth throughout all school districts in the state of North Carolina. This metric, which is strictly quantitative, relies on data about how students have historically performed. Teachers are then assigned a rating based on how students performed on their standardized assessments compared to how they were projected to perform. Principals in my study were asked to give feedback on their perspective of whether or not the NCEES ratings

and the EVAAS scores for teachers are correlated with one another, and also asked to reflect on whether or not they should be.

### **Summary of Findings**

I analyzed the responses provided by principals to the interview protocol. From this analysis, I discerned themes that emerged from the responses of participants. In this study I examined the perspectives of seven (7) principals from the Northeast and North Central regions of North Carolina regarding the correlations between NCEES and EVAAS evaluations of teacher performance. The intent of this research was to determine whether or not principals feel that these two evaluation systems are aligned, if they should be aligned, and what factors may improve the correlations between the two.

In North Carolina teachers are evaluated using both subjective and objective measures of performance. NCEES is considered to be the subjective measure of teacher performance since principals rate teachers themselves using a rubric with this measure. EVAAS is considered to be the objective measure of the teacher evaluation process since EVAAS uses only student testing data to determine the rating for teachers. Many states use similar evaluation tools to evaluate teachers, and historically the correlation between these two measures has been low (Henry & Guthrie, 2015). In this study, principals were interviewed to gain insight into their opinions on such correlations and whether they believed there were particular teacher attributes that would contribute to an increase in student achievement, and inclusion of such attributes in the NCEES rubric might increase the correlation between NCEES and EVAAS ratings.

The research questions that were addressed in this study are as follows:

1. How do the correlations of NCEES and EVAAS ratings compare among the participating districts in the Northeast and North Central regions of North Carolina?



2. How do principals describe the levels of correlation that exist between the district's NCEES and EVAAS ratings?
3. How do principals describe the levels of correlation that exist between their teachers' NCEES ratings and EVAAS ratings and those of the district?
4. To what factors do principals in the Northeast and North Central regions of North Carolina attribute the level of correlation between the NCEES and EVAAS ratings of teachers?
5. What teacher attributes do principals in the Northeast and North Central Regions of North Carolina believe would increase student achievement and thus produce a stronger correlation in NCEES and EVAAS ratings, if an evaluation of these attributes existed in the NCEES?

Research Question 1 focused on comparing the district specific data that were sent to the superintendents of participating districts by the NCDPI. Henry and Guthrie (2015) found EVAAS and NCEES ratings to be loosely correlated in their study of the 2010-2011 through 2013-2014 school year data. In this study the overall correlations between EVAAS and NCEES, as well as the correlations between EVAAS and Standard 4 of the NCEES were not substantially different. Standard 4 is addressed because it is the standard that is most aligned with facilitating instruction (Standard IV: Teachers facilitate learning for their students) (McREL, 2015). A review of the data from participating school districts indicated similar findings. A greater level of correlation existed regarding teachers who were rated at either the upper extreme (Distinguished) or lower extreme (Developing) of the NCEES rubric and the EVAAS scores for those same teachers. Teachers who were rated as Distinguished in NCEES were more likely to be rated as Exceeds Expected Growth in EVAAS. Teachers who were rated as Developing in NCEES were

more likely to be rated as Does Not Meet Expected Growth. It is important to note, however, that the rating of Developing is actually not the lower extreme of the NCEES rubric. Not Demonstrated is actually the lower extreme of this scale, but that rating did not show up in the data that were sent by the NCDPI. This would indicate that no teachers received such a rating in any of the participating districts.

Research Question 2 focused on how participants described the EVAAS and NCEES correlations in their respective school districts. In general, participants expected to see some level of correlation between a teacher's rating on Standard 4 in the NCEES and their EVAAS rating. Five of seven participants felt that the EVAAS and NCEES ratings did not show much correlation at all. One of the seven participants felt that there was some alignment between the upper and lower extremes of NCEES ratings and their corresponding EVAAS ratings. One of the seven participants believed that their district's EVAAS and NCEES ratings correlated well.

Research Question 3 focused on determining how principals described the comparison of their own teachers' ratings in EVAAS and NCEES with those of their district. The participants from this study felt that they were able to accurately evaluate their staff in a manner that would be more aligned with EVAAS than what their district results displayed, but they also conceded that there were times when they would also rate teachers in a way that aligned with a teacher's perspective of themselves so that they could get the evaluation completed without the hindrance of a debate over varying perceptions of levels of performance. The principals reflected on the need for additional documentation that is required if teachers were rated below proficient and the reluctance to give such a rating unless the principal felt that the teacher's poor performance merited their contract not being renewed in the following year. In scenarios like these, participants admitted that these evaluations could skew the correlations between EVAAS and

NCEES, especially on the lower end of both evaluation tools. Some participants also pointed to expectations or mindsets from district level leaders that beginning teachers likely should not be rated at the higher end of the NCEES scale due to lack of experience. This could impact their own correlations between NCEES and EVAAS measures on the upper end of both rubrics if they felt that a beginning teacher deserved a higher rating in NCEES but wanted to ensure that their evaluations aligned with the expectations of those leaders in their respective districts.

Research Question 4 focused on principal perceptions of the factors to which they attribute the limited correlation that exists between NCEES and EVAAS ratings. In their responses participants said that the factors that would contribute to NCEES and EVAAS correlations in their own experience could likely be extrapolated and applied to results across the district and likely the state. Throughout their careers participants in this study had communicated with other school and district level administrators who had had similar experiences. Participant responses indicated that the practice of rating teachers in a manner that would more closely align with their own perception of their abilities rather than going exactly by the rubric in NCEES was rather pervasive. The idea that beginning teachers should not be rated at the upper end of the NCEES scale only came up in three of the seven interviews.

Research Question 5 focused on principal perceptions of teacher attributes that would contribute to an increase in student achievement, and thus, would likely create a greater correlation between NCEES and EVAAS ratings. During the interviews a few common teacher attributes were mentioned by multiple participants. Those common attributes were: a level of skill in providing appropriate levels of questioning in the classroom; teacher expertise in the content standards that they are expected to teach; and the ability for a teacher to develop strong and genuine relationships with their students. Of these, the teacher attribute that was mentioned

most frequently was teacher-student relationships. Five of the seven participants specifically stated that one of the things they look for the most when evaluating a teacher is the type of relationships they have with their students.

### **Discussion of the Study Findings**

To begin my study, the North Carolina Department of Public Instruction sent NCEES and EVAAS correlation data for middle grades mathematics teachers to the superintendents of participating school districts. The superintendents then sent these data to me via email. These data showed that weak correlations existed between these two evaluation metrics in NCEES Standards 1 and 4. To gain deeper insights into the low correlations between NCEES and EVAAS scores, I sought out the perceptions of principals in the Northeast and North Central regions of North Carolina. In order to provide background knowledge on these participants, I analyzed and reported their demographic data. The demographic data that were reported were: race, gender, years of experience, years in their current district, years in their current school, percentage of students at their school who receive free or reduced priced lunches, and whether their school was located in a rural, suburban, or urban area. I developed an interview protocol and conducted conversations with participants using phone or video conference. The information gathered through these quantitative data and the subsequent interviews was used to answer the research questions in the study.

The theoretical framework that undergirded this study was based on three theories that provide insight into the complexities of school leadership and the potential correlations, or lack thereof, between NCEES and EVAAS ratings. Social network theory, cognitive evaluation theory, and expertise theory all provide such insights. Social network theory points to the dynamics that surround human relationships and just how diverse those relationships can be.

This theory would suggest that teachers would be more likely to be receptive to feedback if they feel that the individuals conducting their observations are working alongside them and providing feedback as a colleague, in lieu of coming in as an evaluator who may be making decisions regarding their employment status (Liou et al., 2015). Cognitive evaluation theory posits that perceived competence increases intrinsic motivation. In light of this, providing positive feedback to teachers should increase their confidence and intrinsic motivation regarding their careers (Deci et al., 1999). Expertise theory would suggest that individuals progress along a continuum throughout their years of experience to become experts in their field (Ericsson et al., 1993), and thus, it may not be effective practice to evaluate beginning teachers and more tenured teachers using the same metrics.

### **Discussion of Quantitative Findings**

Research Question 1 focused on the comparison of NCEES and EVAAS correlations among participating school districts. An observation of the data that were sent to participating districts and then forwarded to me revealed a limited correlation between NCEES and EVAAS ratings of teachers in participating districts. There were some instances of stronger correlations at the extremes of the two scales, yet there were also some distinct outliers that were rated at opposite extremes of the two scales.

With NCEES, principals evaluate teachers using a rubric that has been provided by the North Carolina Department of Public Instruction. Teachers can be rated as: not demonstrated, developing, proficient, accomplished, or distinguished. These ratings are based on principals observations and assessments throughout the school year. In this rubric, however, there are 147 possible boxes that can be checked for a teacher's evaluation if they are on a "comprehensive" evaluation cycle. The comprehensive cycle requires that teachers be evaluated four times

throughout the school year. Three of those observations must be conducted by an administrator, and one must be conducted by a peer (McREL, 2015). Teachers on the Comprehensive Cycle are also rated on all five standards of the rubric. Those standards are:

- Standard I – Teachers demonstrate leadership,
- Standard II – Teachers establish a respectful environment for a diverse population of students,
- Standard III – Teachers know the content they teach,
- Standard IV – Teachers facilitate learning for their students, and
- Standard V – Teachers reflect on their practice.

Teachers can also be placed on an “abbreviated” evaluation cycle instead of a comprehensive cycle. Teachers who are on the abbreviated cycle are only evaluated on Standards 1 and 4 in NCEES (McREL, 2015). Since all teachers are rated on Standards 1 and 4, these two standards were chosen as the ones for which I would assess correlations between NCEES and EVAAS measures.

In EVAAS, teachers are provided with a score of does not meet expectation, meets expectation, or exceeds expectation based on how their students perform on standardized tests. For teachers to be rated as does not meet expectation, a substantial amount of student scores that are attributable to that teacher must be falling below their expected score for that particular assessment. The same is true for a teacher to be rated as exceeds expectation, except that a substantial amount of the scores must be surpassing their expected score. For teachers rated as meets expectation, the scores that are attributable to those teachers fell within a range that showed no statistical difference compared to other teachers across the state (SAS Institute, Inc. 2016b).

The weak correlations between NCEES and EVAAS ratings of teachers displayed in the data provided for participating districts aligned with findings in previous research (Henry & Guthrie, 2015). Henry and Guthrie (2015) examined the relationship between and found a low correlation factor of approximately 0.26 for mathematics teachers. While I was not able to calculate such an effect size with the data provided by the NCDPI, an overview of the data that were provided did show a similar trend.

### **Discussion of Qualitative Findings**

Research Question 2 focused on how participants perceived the NCEES and EVAAS correlations from their respective districts. The preliminary interaction for this part of the interview involved each participant reviewing the data described in Research Questions 1 that depicted their district's NCEES and EVAAS correlations. Five out of seven participants acknowledged that the data were essentially a fair depiction of their district with respect to these two metrics. One participant did note that while there was very little correlation overall, the scores on the upper ends of both scales were more closely aligned (specifically the "accomplished" section in NCEES and the "meets" or "exceeds expectation" sections in EVAAS. Another participant stated that he felt that the NCEES and EVAAS scores for his district were well aligned with one another. It is important to note here that some of the participating school districts were smaller districts and had very few middle schools in their district. With that in mind, it is possible that participants in those districts may have felt more personally responsible for the correlations, especially if there was only one middle school in the district. This may have prompted the latter two participants to be more prone to defend their district's results.

While I was unable to locate any other studies that specifically asked principals to reflect on correlations between their own district specific measures of teacher effectiveness, other studies do exist that align with the responses of participants in this study. The notion that correlations between the two evaluation systems are weakly correlated except at the extremes aligns with the findings of Harris et al. (2014). In their study a stronger correlation was found at the high end of both evaluation measures, yet weak correlations existed elsewhere in the scales. Additional studies from the body of research on this topic also point out the concerns with the lack of correlation that historically exists between these two types of teacher evaluations. Some explicitly address the question of whether one should even assert that these two measures should be correlated in the first place, as they are likely measuring different aspects of teaching (Amrein-Beardsley & Collins 2012; Ballou & Springer, 2015; Baker et al. 2010). While there are certainly a variety of metrics that could be used to evaluate teacher performance, some connection should exist between what a principal sees in their classroom and their impact on student growth as measured by EVAAS. Teaching is a dynamic profession and should be evaluated in such a manner, but it is imperative that teachers be given feedback that would assist them in helping their students make academic gains. Such feedback is limited in its utility if correlations do not exist between objective measures of performance like EVAAS and elements of what can be seen and subjectively evaluated in the instructional process.

Research Question 3 focused on how participants described their own evaluations of teachers. Participants were then asked to reflect on how those evaluations would correlate with EVAAS and to reflect on those evaluations in light of the district-specific correlations that were provided by the NCDPI. In their interviews, participants reflected that they felt knowledgeable enough about instruction and performance appraisal to accurately gauge a teacher's likely impact



on student achievement. However, participants also discussed the impact of teacher shortages and requirements around next steps when teachers are rated below proficient. They expounded upon how those factors can impact their end-of-year ratings for teachers. Teacher shortages factor into their considerations when rating teachers simply because it has been difficult, at times, in recent years to fill teaching positions. If teachers are unhappy with their ratings in NCEES, they may feel unappreciated and look for employment elsewhere. This would leave the principal in a position of searching for a new teacher for the next school year.

Participants also commented on the obligations placed on them by their respective districts that require them to take extra steps when teachers are deemed to be performing unsatisfactorily. If a teacher is rated as anything below Proficient in their overall summative rating, the administrator is expected to place that teacher on a performance plan and track their progress in a detailed manner the following school year. This information is then used as a potential justification for a teacher's contract to not be renewed. Considering such obligations, it is possible that even if a principal feels that a teacher is underperforming, they might decide to inflate the evaluation and provide a more positive rating than is warranted. Unless the administrator feels so strongly about the matter that they feel the teacher's contract should be non-renewed, they may feel inclined to adjust their ratings in order to avoid the process of placing a teacher on a performance plan. A recent study found that principals may be inclined to change a teacher's grade level or content area to an untested subject if they deem them as underperforming instead of having to place them on an improvement plan (Goldring et al., 2015).

Although there is limited research that has asked principals to reflect on the correlations among elements of their own teacher evaluations, the concept of principals potentially inflating

teacher ratings for various reasons has been documented in a recent study by Jones et al. (2022). These authors found that a myriad of reasons that principals might elect to be more lenient in their evaluations. Two of the incentives presented in their study include “managing limited time wisely” (p. 60) and “avoiding difficult conversations” (p. 61). These findings align well with the reflections of the participants in my study. Several participants in my study mentioned similar concerns when conducting teacher evaluations. They knew that if a teacher were to be rated below “Proficient” in NCEES, then the following year they would need to have that teacher on an action plan and follow up with them routinely to ensure that they were making progress on that plan, according to their district requirements. That would require a substantial amount of time on their part when their time is already limited. Drake et al. (2015) reached similar conclusions in their study of six urban school districts where they conducted 56 principal interviews. Principals in those districts also felt that the process for working with teachers on improvement plans was time consuming.

With regard to avoiding difficult conversations, participants also mentioned that unless a teacher is performing so poorly that they are considering not renewing their contract, such teachers are often given the benefit of the doubt and rated in a manner that would likely be more acceptable to the teacher. Participants mentioned being tasked with making sure that they have vacancies filled and preferably filled with high quality teachers. Given the risk that a teacher might have decided to leave if they were unhappy, it sometimes made more sense for keep teachers in place, and sometimes that could mean making sure that the teacher feels comfortable with their ratings in NCEES.

Research Question 4 focused on principal perceptions of factors to which they attributed the level of correlation that exists between NCEES and EVAAS ratings. As participants

described their lived experiences, they also commented on the idea that the factors that impact their evaluations of their own teachers are those that likely impact other principals across the district and state. From their conversations with other teachers and administrators during their tenure in education, the concerns among the participants are also being felt by others. Some of the factors they discussed were mentioned in Research Question 3 (e.g., principals being willing to be more lenient in their evaluations for various reasons), but a few additional themes emerged, as well.

**Interventions that were not provided by the teacher of record.** Multiple participants mentioned the potential for at least some of the lack of correlation between NCEES and EVAAS ratings to be attributable to district and school level interventions that were in place to impact student achievement. For example, schools that receive Elementary and Secondary School Emergency Relief (ESSER) funds and/or Title I funds often pay for tutors with some of those funds. The tutoring that students receive may not necessarily be from the teacher whose EVAAS rating would be calculated using those students' test scores. Such interventions would likely impact correlations between a teacher's rating in NCEES and their correlated EVAAS rating.

**Lack of aligned ratings among evaluators.** During their interviews, multiple participants expressed that a lack of alignment exists among evaluators who observe the same teacher. This points to a lack of consistent training that could be provided by either the state or district level. If different evaluators go into a teacher's classroom with different expectations of what constitutes a "proficient," "accomplished," or "distinguished" rating, then there will likely be fluctuations in teacher ratings.

**NCEES not aligned with EVAAS and student achievement.** Multiple participants in this study indicated that they did not feel that NCEES adequately addressed some of the factors

in a school or classroom that they believed contribute to an increase in student achievement. Some of those factors are the types of questions teachers are using in the classroom; the opportunities for students to collaborate on assignments; a teacher's knowledge of the content standards; various supports that are put in place for students; classroom management; and how a teacher interacts with their students both inside and outside of the classroom. Participants felt that how a school functions and the dynamic manner in which adjustments are made throughout the school year may be difficult to incorporate into a teacher evaluation, but they did feel that these elements do impact student achievement.

Much of the literature around this topic focuses on which type of evaluation method, subjective or objective, is better. Limited literature exists on whether NCEES is actually measuring the correct things or if districts and the state have a common language around NCEES metrics. The findings in my study, therefore, appear to be somewhat novel. It is important to point out a reflection from one participant in particular. In her interview, she contended that even if there are major overhauls in the standards that are addressed in NCEES, these evaluations are still going to be based on the opinions of others and will thus yield themselves to at least some degree of subjectivity, thus impairing the degree of alignment among the performance metrics. Specifically she stated "...it's still going to come down to what the observer does at that time to rate [the teacher] on that [standard]...students and growth in the school is still going to come down to someone else's observation...."

Research Question 5 focused on principal perceptions of teacher attributes that would contribute to an increase in student achievement, and thus would likely create a greater correlation between NCEES and EVAAS ratings. In their interviews, participants provided insights into attributes that they see in teachers who traditionally have positive impact on student

achievement. From their responses, two themes emerged: 1) relationships and 2) pedagogy. Five out of seven participants mentioned relationships with students directly as one of the main factors that they believe contribute to an increase in student achievement. The second theme that emerged was teacher pedagogy. Within that theme, there were three subthemes: 1) the ability of teachers to align their instruction to the North Carolina Standard Course of Study (NCSCOS), 2) how teachers engage in questioning in the classroom, and 3) whether or not students are actively engaged in the classroom.

The themes that emerged from participant responses in my study align well with recent research that was published by Baier et al. (2019). In this study 209 German mathematics teachers were studied and their methods for successful increases in student achievement were ascertained. The 209 teachers served 4,672 students who ranged in ages from 13 to 16 years old. The researchers found that one of the attributes of a successful mathematics teacher was that they were rated highly with respect to their "caring ethos." This would align with the relationship element that was mentioned by several participants in my study. Two other elements that were listed in the Baier et al. study were "constructive response to errors" and "discursive treatment," which align with the responses in my study that were deemed as subthemes within the broader theme of pedagogy.

### **Reflections on Theoretical Framework**

The three theories that provided a foundation for my research were social network theory, cognitive evaluation theory, and expertise theory. Social network theory indicates that when individuals feel that they are working together alongside their supervisors or superiors, they are more likely to be receptive to feedback. This would suggest that if the personnel who are conducting evaluations of teachers are seen in the role of a coach instead of someone who is

making a decision regarding their continued employment, teachers would be more likely to receive this feedback in a positive manner and thus more likely to implement said feedback. Cognitive evaluation theory suggests that positive feedback increases intrinsic motivation in employees. From my experience as an educator, there exists a lack of positive feedback in NCEES evaluations. This would indicate that teachers may view NCEES evaluations in a negative manner if they are not receiving specific positive feedback within this evaluation tool. Expertise theory would suggest that beginning teachers and veteran teachers would likely exhibit different strategies and behaviors in a classroom, and thus should be evaluated using different metrics. It is my belief that the perceptions of principals will align with this theory and that they will have different indicators that look for when determining if a beginning teacher is having a successful school year, and if a veteran teacher is doing the same.

These three theories provided a structure around which I based my research questions and interview protocol. They also offered a framework that guided me in reviewing the interview transcripts while conducting in-vivo coding. When analyzing and summarizing the results of my study, these theories also provided scaffolding around which I could build my results.

All three of the theories that made up my theoretical framework (social network theory, cognitive evaluation theory and expertise theory) were well-aligned with the results of my study. Participants indicated that there is a dynamic to the evaluation process for teachers that involves how receptive teachers are to evaluations that feel more formative in nature versus evaluations that feel summative such as NCEES. This is in line with what social network theory suggests. From the responses provided by participants, it is also evident that positive feedback plays a major role in whether or not teachers feel receptive to their evaluation feedback; this is consistent with the elements of cognitive evaluation theory. Participants also discussed mindsets from

leaders in education that indicated that beginning teachers should not be rated as highly as veteran teachers on the NCEES scale due to their lack of experience. Such mindsets align with expertise theory.

### **Limitations of the Study**

One of the limitations of this study was that the interviews were conducted during the COVID-19 pandemic. That caused these interviews to have to be conducted through telephone or video conferencing. There is a chance that conducting the interviews in this way could have made participants either more or less likely to share information depending on their comfort level and trust in providing responses in such a setting. The COVID-19 pandemic may have also inadvertently limited the sample size of this study. At the time when I was requesting permission from school districts to conduct my study, it is likely that superintendents were receiving a multitude of emails; this may have impacted whether or not they even saw my request to conduct this research in their district.

Another limitation of the study is the fluctuation in the number of observations a teacher received during the school year based on whether they were on a comprehensive cycle, a standard cycle, or an abbreviated cycle. When some teachers are observed more than others before receiving a final rating in NCEES, the correlations of those ratings could be impacted.

An additional limitation of this study was the size of some of the participating school districts. With some of the districts only having one or two middle schools, the districtwide results of the NCEES and EVAAS correlations were likely much more personal to the participants in school systems. With the small size of their school district, these principals were aware that the correlations, or lack thereof, in their district's results were largely attributable to them.

One final limitation of this study was the lack of training that had been provided to participants on either evaluation tool, as self-reported by the participants. Six of the seven participants in this study reported that the school where they served was located in a rural setting, and six of the seven participants reported that approximately 60% or more of their students received free or reduced priced lunches. Often schools that are located in high-poverty and rural areas do not have as much funding available as some of their more affluent counterparts. Inadequate funding can often result in diminished access to training; this could have impacted some of the responses of participants as it relates to their understanding of the evaluation tools.

### **Implications of the Study for Policy and Practice**

This study took place in districts in the Northeast and North Central regions of North Carolina. Based on quantitative NCEES and EVAAS data provided by NCDPI for the participants' districts, the current study found that the ratings of middle grades mathematics teachers on the NCEES and EVAAS scales were weakly correlated. There was not a distinguishable difference between the correlations of Standard 1 and Standard 4 of NCEES when compared with EVAAS ratings. The interviews with middle school principals were aligned to these patterns in the state-provided data. The principals largely felt that the NCEES and EVAAS correlations from their respective districts for middle grades mathematics teachers were weakly correlated. When discussing why these two evaluation measures are not more closely correlated, participants provided insights from their backgrounds and experiences in the field of education. Our discussions disclosed that these participants received very little training on the evaluation measurements that are used to judge teachers in the state of North Carolina. If principals are expected to provide teachers with accurate evaluations and feedback, then there should be a consistent effort to ensure that adequate training is provided to all stakeholders.



## **Implications for Policy**

Accountability will always need to be a factor in public education. Public schools are funded with public dollars, so a level of oversight is owed to the public to ensure an adequate education is provided to the students who are served in our schools (Danielson, 2012). Those who serve our students, however, also need to know and understand the measures by which they are being evaluated in order to be held accountable. They have to be familiar with those measures in order for any level of accountability to be successful. Based on the participant responses in this study, I make the following recommendations for boards of education at the state and district levels:

- Ensure that teachers and supervising administrators at all levels are provided with professional development on both the NCEES and EVAAS evaluation tools;
- Ensure that schools and districts located in high-poverty areas have the same access to professional development on these evaluation tools as other school districts;
- Ensure that the superintendent and district level leaders are able to adequately provide support to both principals and teachers on how to use the evaluation tools in a formative manner;
- Create calendars that ensure NCEES evaluations provide a framework for implementing evaluations in a timely fashion;
- Require that formative feedback be provided to teachers in addition to the formal rubric ratings and establish training for such evaluative processes;
- Ensure that aggregate NCEES and EVAAS ratings are reviewed at the state and district levels at least annually;

- At the state level, add in an area to NCEES that would allow principals to predict what they believe a teacher's EVAAS rating will be based on the observations that they are conducting and an area to provide feedback on why they would make that prediction. The responses in this area could then be used to find common themes happening around the school, district, and state regarding what is going well in classrooms and what areas of concern exist. State level leaders can then provide support to regions and/or districts based on the needs that are suggested by such data.
- If the ability for a principal to predict a teacher's EVAAS rating were integrated into NCEES in a way that only the macro-level ratings would be shown to the district and state level leadership, principals would likely be more willing to provide accurate ratings for their teachers. If teachers are not able to see what a principal's prediction is, then there is less of an incentive for principals to inflate those ratings, especial if the data are only used in a formative capacity. Once themes can be developed from the data that that are entered, the NCDPI could then adjust Standard 4 in NCEES and create a new standard that would be more aligned with evaluating the attributes of teachers who consistently produce increases in student achievement.

### **Implications for Practice**

Providing feedback to both teachers and administrators is one of the most important elements ensuring that educators perform in a manner that enables public school students to receive an appropriate education. If evaluations are conducted and worthwhile, feedback is not given, then the value of the evaluation is lost. Both the NCEES and EVAAS evaluation tools provide valuable information and data that can be used to support teachers, principals, and district level leaders. Therein lies the area of importance; the information and data have to be

used. Educators should be willing to act on the results of these evaluation measures to make sure that teachers and administrators are consistently strengthening their abilities (Fusarelli, 2008). I make the following practice-oriented recommendations to school, district, and state level leaders in education to ensure that evaluations are being used in such a way:

- District leaders should monitor adherence to the above-mentioned calendars to ensure NCEES evaluations are conducted in a timely fashion;
- District leaders should place more emphasis on teacher growth and progression rather than overall ratings by reviewing the feedback that was provided to teachers and monitoring the steps that they took to implement that feedback in the classroom;
- District leaders and principals should have teachers who are strong in certain areas provide professional development and support to their colleagues who may be struggling in that same area. This recommendation arises from the work of Borko et al. (2021) that I reviewed for this study;
- District leaders should provide quarterly and yearly training on what administrators should be looking for when they enter a teacher's classroom if a teacher is to be rated at certain levels on the NCEES rubric;
- District leaders should create opportunities for collaboration among administrators during the summer months at a minimum, quarterly preferred, so that teacher performance can be reviewed and administrators can support one another in helping teachers across the district and state;
- As a local dimension of the policy recommendation that NCEES include a section in which principals predict a teacher's EVAAS rating, district leaders should provide support to specific schools and principals based on the themes that arise from the

related analyses. Principals can use this same information to provide support at the school level based on the themes they are seeing within their own buildings. In addition to proper support being provided, this information can be used for accountability measures as well. If principals are predicting that a teacher's results will be poor in EVAAS, then they should be providing support to that teacher on how to get better. If districts are seeing that multiple schools are having concerns in the same area, then district level leaders should be held accountable for providing support in those areas.

### **Recommendations for Future Research**

My study can serve as a framework for future research on the topic of correlations between subjective and objective measures of teacher evaluation. My study focused on principal perceptions in the Northeast and North Central regions of North Carolina who served schools where greater than 50% of students received free or reduced priced lunches. Future researchers should consider conducting this research in other regions of North Carolina and with principals who serve schools that represent a broader range of socioeconomic profiles. Six of the seven participants in my study also served in schools that they self-reported as rural. Future researchers should consider gaining insights from principals in both suburban and urban areas.

Another topic for future researchers to consider would be to have principals predict where they believe teachers will be rated in EVAAS and compare that against the NCEES ratings for those same teachers to see if correlations exist in that regard. These predictions could be conducted at various points in the school year and then compared against actual EVAAS ratings for those same teachers once those results are released. Determining whether principals

have an accurate estimation of a teacher's impact on student achievement would be quite valuable and serve as a worthwhile foundation for related professional development.

Future researchers should also consider conducting similar studies with educational stakeholders other than principals. A similar study that obtains feedback from teachers, district level leadership, and board members at both the district and state levels would provide insights that are almost nonexistent in the contemporary body of educational research.

Future researchers should also consider verifying the data literacy of participants in the study prior to conducting interviews. Some of the graphs and data analysis could be difficult for participants to understand and thus an evaluation of their data literacy would provide valuable information to the researcher before moving forward with the study.

### **Chapter Summary**

Accountability in public education can often be a polarizing topic of discussion. Public schools are funded by public money, and thus have a responsibility to the public to ensure that children are receiving a proper education (Danielson, 2012). Teacher evaluations are one way that such accountability is implemented. In North Carolina teachers are evaluated using both a subjective measure called the North Carolina Educational Evaluation System (NCEES) and an objective measure called the Education Value-Added Assessment System (EVAAS). In this study I examined the perceptions of middle grades principals about the correlations between those measures.

The participants in this study were located in the Northeast and North Central regions of North Carolina and six of these seven principals worked at schools where greater than 50% of the students received free or reduced priced lunches. Participants were provided with graphs from the North Carolina Department of Public Instruction (NCDPI) that displayed their districts'

correlations between ratings for middle grades mathematics teachers on Standards 1 and 4 in NCEES and their districts' ratings on EVAAS for those same teachers. For both Standard 1 and 4, weak correlations existed between the NCEES ratings and EVAAS ratings of those same teachers. Correlations were slightly stronger towards the low and high extremes of both scales.

In the review of archival data, this study found weak correlations between NCEES and EVAAS ratings of middle grades mathematics teachers in participating districts. These results align with findings from similar studies conducted in recent years (Batten et al., 2012; Henry & Guthrie, 2015). Each participant reviewed these data at the beginning of a conversation in which I used an original interview protocol to solicit their feedback on these correlations and their perceptions about such correlations. This study found that participants had received very little training on either evaluation method in recent years and that a consistent understanding of how to implement these measures did not exist among the participants. Some of the leaders in the districts where these principals served had more conversations with participating principals about how alignment should exist between the two evaluation measures, but there was a lack of concrete examples regarding how to make sure that such alignment happened.

Based on these results, it is recommended that policymakers and leaders in education should reevaluate the two measures by which teachers are evaluated in North Carolina to see if adjustments can be made in order to create more alignment between these two systems. Policymakers could also evaluate the use of EVAAS ratings as they pertain to teacher ratings. The students of North Carolina deserve well-informed teachers who can provide them with strong potential for academic gains (Hostetler et al., 2007).

The study concluded with potential implications for policy and practice based on the results of this study. Suggestions for future research are also provided.

## REFERENCES

- AERA Council. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452.
- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75.
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value-Added Assessment System (SAS® EVAAS®) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives EPAA*, 20(12-42). Retrieved Feb. 18, 2016, from <http://epaa.asu.edu/ojs/article/view/1096>
- Amrein-Beardsley, A., Pivovarova, M., & Geiger, T.J. (2016). Value-added models: What the experts say. *Phi Delta Kappan*, 98(2), 35-40.
- Anderson, N. (2013). D.C. school officials: 44 teachers were given mistaken performance evaluations. *The Washington Post*. Retrieved April 3, 2017, from [https://www.washingtonpost.com/local/education/dc-school-officials-44-teachers-were-given-mistaken-performance-evaluations/2013/12/23/c5cb9f26-6c0c-11e3-a523-fe73f0ff6b8d\\_story.html?utm\\_term=.ac4add25a18e](https://www.washingtonpost.com/local/education/dc-school-officials-44-teachers-were-given-mistaken-performance-evaluations/2013/12/23/c5cb9f26-6c0c-11e3-a523-fe73f0ff6b8d_story.html?utm_term=.ac4add25a18e)
- Archer, J., Cantrell, S., Holtzman, S.L., Joe, J.N., Tocci, C.M., & Wood, J. (2016). *Better feedback for better teaching: A practical guide to improving classroom observations*. Jossey-Bass. Retrieved March 26, 2017, from <http://k12education.gatesfoundation.org/wp-content/uploads/2016/05/BetterFeedbackBetterTeaching.pdf>

- Baier, F., Decker, A.-T., Voss, T., Kleickmann, T., Klusmann, U., & Kunter, M. (2019), What makes a good teacher? The relative importance of mathematics teachers' cognitive ability, personality, knowledge, beliefs, and motivation for instructional quality. *British Journal of Educational Psychology*, 89, 767-786. <https://doi-org.prox.lib.ncsu.edu/10.1111/bjep.12256>
- Baker, E.L., Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., Ravitch, D., Rothstein, R., Shavelson, R.J., & Shepard, L.A. (2010). Problems with the use of student test scores to evaluate teachers. *Economic Policy Institute*, Briefing Paper #278.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *American Educational Research Association and American Statistical Association*, 29(1), 37-65.
- Ballou, D., & Springer, M. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44(2), 77-86.
- Bhandari, P. (2023). Triangulation in research: Guide, types, examples. Scribbr. Retrieved from <https://www.scribbr.com/methodology/triangulation/>
- Batten, D., Britt, C., DeNeal, J., & Hales, L. (2012). NC teacher evaluations & teacher effectiveness: Exploring the relationship between value-added data and teacher evaluations (Rep.). Retrieved from <http://www.dpi.state.nc.us/docs/intern-research/reports/teachereval.pdf>
- Bleiberg, J., Brunner, E., Harbatkin, E., Kraft, M., & Springer, M. (2024). Taking teacher evaluation to scale: The effect of state reforms on achievement and attainment.



- (EdWorkingPaper: 21-496). Retrieved from Annenberg Institute at Brown University:  
<https://doi.org/10.26300/b1ak-r251>
- Burns, A.C. (2006). Teaching experientially with the Madeline Hunter method: An application in a marketing research course. *Simulation & Gaming, 37*(2), 284-294.
- Butcher, K.F., & Visher, M.G. (2013). The impact of a classroom-based guidance program on student performance in community college math classes. *Educational Evaluation and Policy Analysis, 35*(3), 298-323.
- Borko, H., Carlson, J., Deutscher, R., Boles, K., Delaney, V., Fong, A., Jarry-Shore, M., Malamut, J., Million, S., Mozenter, S., & Villa, A.M. Learning to Lead: an Approach to Mathematics Teacher Leader Development. *International Journal of Science and Mathematics Education 19*(Suppl 1), 121–143. <https://doi-org.prox.lib.ncsu.edu/10.1007/s10763-021-10157-2>.
- Chiang, H., Speroni, C., Herrmann, M., Hallgren, K., Burkander, P., Wellington, A. (2017). Evaluation of the teacher incentive fund: Final report on implementation and impacts of pay-for-performance across four years. National Center for Education Evaluation and Regional Assistance. Retrieved from <https://ies.ed.gov/ncee/pubs/20184004/pdf/20184004.pdf>
- Cohen, D., Raudenbush, S., & Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119-142. Retrieved from <http://www.jstor.org.prox.lib.ncsu.edu/stable/3699546>
- Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS education value-added assessment system (EVAAS®). *Education Policy Analysis Archives, 22*(98), 1-42. <http://dx.doi.org/10.14507/epaa.v22.1594>

- Condie, S., Lefgren, L., & Sims, D. (2014). Teacher heterogeneity, value-added and education policy. *Economics of Education Review*, 40, 76-92.
- Creswell, J.W. (2014). *Research design* (4<sup>th</sup> ed.). SAGE Publications, Inc.
- Cubberley, E.P. (1929). *Public school administration* (3rd ed.). Houghton Mifflin.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Association for Supervision and Curriculum Development.
- Danielson, C. (2012). It's your evaluation— Collaborating to improve teacher practice. *Education Digest*, 77(8), 22-27. Retrieved April 11, 2016.
- Danielson, C. (2016). National Association of State Boards of Education (NASBE) 2016 Legislative Conference, April 5, 2016.
- Darling-Hammond, L. (2012). Value-added evaluation hurts teaching. *Education Week*, 31(24), 24, 32.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15. Retrieved from <http://www.jstor.org/stable/41497541>
- Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. Teachers College Press.
- Davis, C.R., Bangert, L., Comperatore, A.N., & Smalenberger, M. (2015). Teacher and principal perceptions of the North Carolina educator evaluation system. *Consortium for Educational Research and Evaluation – North Carolina*. Retrieved March 30, 2017, from <https://publicpolicy.unc.edu/files/2015/07/Teacher-and-Principal-Perceptions-of-the-North-Carolina-Educator-Evaluation-System-September-2015.pdf>

- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627-668. doi:10.1037/0033-2909.125.6.627
- Dowley, R.G., & Kaplan, N. (2014). Evaluating evaluation: Assessing Massachusetts school districts' implementation of educator evaluation requirements. *Journal of Law and Education*, *43*(4), 485-502.
- Drake, T.A., Goldring, E.B., Grissom, J.A., Cannata, M.A., Neumerski, C.M., Rubin, M. & Schuermann, P. (2015). Development or dismissal: Exploring principals use of teacher effectiveness data. In Grissom, J.A. and Youngs, P., *Making the Most of Multiple Measures* (pp. 199-214). New York: Teachers College Press.
- Duckor, B., Castellano, K.E., Tellez, K., Wihardini, D., & Wilson, M. (2014). Examining the internal structure evidence for the performance assessment for California teachers: A validation study of the elementary literacy teaching event for Tier I Teacher Licensure. *Journal of Teacher Education*, *65*(5), 402-420.
- Editorial Projects in Education Research Center. (2015). Issues a-z: No child left behind: An overview. *Education Week*. Retrieved April 2, 2017, from <http://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html/>
- Ericsson, K.A., Krampe, R.T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363-406.  
doi:10.1037/0033-295x.100.3.363
- Everson, K.C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? *Review of Educational Research*, *87*(1), 35-70.

- Excellent Schools Act of 1997. S. L 1997-221. Retrieved March 18, 2017 from <http://www.ncleg.net/EnactedLegislation/SessionLaws/PDF/1997-1998/SL1997-221.pdf>
- Franco, M.S., & Seidel, K. (2012). Evidence for the need to more closely examine school effects in value-added modeling and related accountability policies. *Education and Urban Society, 46*(1), 30-58.
- Fusarelli, L.D. (2008). Flying (partially) blind: School leaders' use of research in decision making. *The Phi Delta Kappan, 89*(5), 365–368. <http://www.jstor.org/stable/20442501>
- Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education, 65*(5), 389-401.
- Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. Holt, Rinehart & Winston.
- Goldring, E., Grissom, J.A., Rubin, M., Neumerski, C.M., Cannata, M.A., Drake, T.A., & Scheurmann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher 44*(2):96-104.
- Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher, 43*(6), 293-303.
- Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice, 15*(2), 273-289.
- Guarino, C.M., Reckase, M.D., & Wooldridge, J.M. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy, 10*(1), 117-156.

- Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development*, 78(1), 116–131.  
[https://doi.org/ 10.1111/j.1467-8624.2007.00988.x](https://doi.org/10.1111/j.1467-8624.2007.00988.x)
- Harris, D.N. (2009). Would accountability based on teacher value-added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy*, 4(4), 319-350.
- Harris, D.N., & Sass, T.R. (2009). What makes a good teacher and who can tell? The Urban Institute. Washington, D.C. Retrieved March 25, 2017, from  
<http://www.urban.org/sites/default/files/publication/33276/1001431-What-Makes-for-a-Good-Teacher-and-Who-Can-Tell-.PDF>
- Harris, D.N. (2011). Value-added measures and the future of educational accountability. *Science*, 333(6044), 826-827. Retrieved from <http://science.sciencemag.org/content/333/6044/826>
- Harris, D.N., Ingle, W.K., & Rutledge, S.A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, 51(1), 73-112.
- Harris, D.N., & Sass, T.R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review*, 40, 183-204.
- Hawkins, B. (2019). Progress, policy, and protests: Teacher evaluation laws evolving faster than underlying research that proves their worth, experts say. Educator Writer’s Association. Retrieved from <https://www.ewa.org/latest-news/progress-policy-andprotests-teacher-evaluation-laws-evolving-faster-underlying-research>

- Heneman III, H.G., Milanowski, A., Kimball, S.M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay*. CPRE Policy Briefs. Retrieved from: [http://repository.upenn.edu/cpre\\_policybriefs/33](http://repository.upenn.edu/cpre_policybriefs/33)
- Henry, G.T., & Guthrie, J.E. (2015). An evaluation of the North Carolina educator evaluation system and the student achievement growth standard: 2010-11 through 2013-14. *Consortium for Educational Research and Evaluation–North Carolina*. Retrieved from: <https://www-data.fi.ncsu.edu/wp-content/uploads/2021/10/26122948/An-Evaluation-of-the-North-Carolina-Educator-Evaluation-System-and-the-Student-Achievement-Growth-Standard.pdf>
- Hewitt, K.K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, 23(76), 76-124. <http://dx.doi.org/10.14507/epaa.v23.1968>
- Hill, H.C., Kapitula, L., & Umland, K. (2010). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.
- Hinchey, P.H. (2010). *Getting teacher assessment right: What policy makers can learn from research*. Boulder, CO: National Education Policy Center. Retrieved March 27, 2017, from [http://nepc.colorado.edu/files/PB-TEval-Hinchey\\_0.pdf](http://nepc.colorado.edu/files/PB-TEval-Hinchey_0.pdf)
- Holdzkom, D. (1991). Teacher performance appraisal in North Carolina: Preferences and practices. *Phi Delta Kappan*, 72(10), 782-785.
- Hostetler, K., Macintyre Latta, M.A., & Sarroub, L. K. (2007). Retrieving meaning in education: The question of being. *Journal of Teacher Education*, 58(3), 231-244.
- House Bill 322. General Assembly of North Carolina Session 2017. Retrieved March 29, 2017, from <http://www.ncleg.net/Sessions/2017/Bills/House/PDF/H322v1.pdf>

- Hunter, M. (1984). Knowing, teaching, and supervising. In P. Hosford (Ed.), *Using what we know about teaching* (pp. 169–192). Association for Supervision and Curriculum Development.
- Hushman, G., & Hushman, C. (2015). Value-added modeling in physical education. *Strategies*, 28(6), 23-27.
- Husserl, E. (2012). *Ideas: General introduction to pure phenomenology*. Taylor & Francis Group. Retrieved November 11, 2023, from *ProQuest Ebook Central*, <https://www.proquest.com/legacydocview/EBC/958202?accountid=12725>.
- Jacob, B., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
- Jennings, D., Surgenor, P., & McMahon, T. (2013). Education Theory/Constructivism and Social Constructivism in the Classroom – UD. Retrieved from [http://www.ucdoer.ie/index.php/Education\\_Theory/Constructivism\\_and\\_Social\\_Constructivism\\_in\\_the\\_Classroom](http://www.ucdoer.ie/index.php/Education_Theory/Constructivism_and_Social_Constructivism_in_the_Classroom)
- Jones, E., Bergin, C., & Murphy, B. (2022). Principals may inflate teacher evaluation scores to achieve important goals. *Educational Assessment, Evaluation and Accountability*, 34(1), 57-88. Retrieved from <https://doi.org/10.1007/s11092-021-09366-8>.
- Johnson, B., & Christensen, L. B. (2000). *Educational Research: Quantitative and Qualitative Approaches*. Boston, MA: Allyn and Bacon.
- Karl, A.T., Yang, Y., & Lohr, S.L. (2013). A correlated random effects model for nonignorable missing data in value-added assessment of teacher effects. *Journal of Educational and Behavioral Statistics*, 38(6), 577-603.

- Kersten, T. A., & Israel, M. S. (2005). Teacher evaluation: Principals' insights and suggestions for improvement. *Planning and Changing* 36(1/2), 47-67. Retrieved from <https://files.eric.ed.gov/fulltext/EJ737642.pdf>.
- Koedel, C., Mihaly, K., & Rockoff, J.E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180-195.
- Lewis, W.D., & Fusarelli, L.D. (2010). Leading schools in an era of change: Toward a “new” culture of accountability? In S. D. Horsford (Ed.), *New perspectives in educational leadership: Exploring social, political, and community contexts and meaning* (pp. 111-125). Peter Lang.
- Lewis, H., & Leps, J.M. (1946). When principals supervise. *Educational Leadership*, 3(4), 160–163.
- Lincove, J.A., Osborne, C., Dillon, A., & Mills, N. (2014). The politics and statistics of value-added modeling for accountability of teacher preparation programs. *Journal of Teacher Education*, 65(1), 24-38.
- Liou, Y., Daly, A.J., Brown, C., & Fresno, M.D. (2015). Foregrounding the role of relationships in reform. *International Journal of Educational Management*, 29(7), 819-837.  
doi:10.1108/ijem-05-2015-0063
- Manning, J. (2017). In vivo coding. In Matthes, J. (Ed.), *The international encyclopedia of communication research methods*. New York, NY: Wiley-Blackwell. Retrieved from <https://doi.org/10.1002/9781118901731.iecrm0270>
- Marchant, G.J., David, K.A., Rodgers, D., & German, R.L. (2015). State teacher evaluation and teacher education. *The Teacher Educator*, 50(2), 89-108.



- Marks, J.R., Fuller, S.C., Guthrie, J.E., Henry, G.T., & Stallings, D.T. (2015). *North Carolina race to the top: Overall impact and implementation findings*. Consortium for Educational Research and Evaluation – North Carolina. Retrieved March 29, 2017, from <https://publicpolicy.unc.edu/files/2015/12/North-Carolina-Race-to-the-Top-Overall-Impact-and-Implementation-Findings-Final-Report-September-2015.pdf>
- Marzano, R.J. (2003). *Building background knowledge for academic achievement*. ASCD.
- Marzano, R.J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. ASCD.
- McREL. (2015). North Carolina teacher evaluation system. Retrieved March 4, 2016, from <http://www.dpi.state.nc.us/docs/effectiveness-model/ncees/instruments/teach-eval-manual.pdf>
- Miller, D.T., & Hanna, R. (2014). *Four years later, are race to the top states on track?* Center for American Progress. Retrieved March 30, 2017, from <https://www.americanprogress.org/issues/education/reports/2014/03/24/86197/four-years-later-are-race-to-the-top-states-on-track/>
- Milner, J. (1991). Suppositional style and teacher evaluation. *The Phi Delta Kappan* 72(6), 464-467.
- Newton, X.A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and context. *Education Policy Analysis Archives*, 18(23), 23-49. Retrieved March 24, 2017, from <http://epaa.asu.edu/ojs/article/view/810>
- No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002)

North Carolina Association of School Administrators. (2017). A-F school performance grades.

NCASA Legislative Link. Retrieved March 29, 2017, from

<http://www.ncasalegislativelink.org/a-f-school-performance-grades.html>

North Carolina Department of Public Instruction. (2015a). 2014 Ready accountability

background brief supplement: North Carolina school performance grades.

Communication and Information Division. Retrieved March 29, 2017, from

<http://www.ncpublicschools.org/docs/accountability/reporting/spgbckgrndpack15.pdf>

North Carolina Department of Public Instruction. (2015b). Licensure section. Retrieved from

<http://www.ncpublicschools.org/src/guide/personnel/>

North Carolina Department of Public Instruction. (2016a). Evaluation instruments and processes.

Retrieved March 4, 2016, from <http://www.dpi.state.nc.us/effectiveness->

[model/ncees/instruments/](http://www.dpi.state.nc.us/effectiveness-model/ncees/instruments/)

North Carolina Department of Public Instruction. (2016b). EVAAS. Retrieved March 4, 2016,

from <http://www.dpi.state.nc.us/effectiveness-model/evaas/>

North Carolina Department of Public Instruction. (2016c). Selection of EVAAS. Retrieved April

11, 2016, from <http://www.dpi.state.nc.us/effectiveness-model/evaas/selection/>

North Carolina Department of Public Instruction (2017a). Public school employee evaluation.

Retrieved March 18, 2017 from <http://www.ncpublicschools.org/district->

[humanresources/evaluation/](http://www.ncpublicschools.org/district-humanresources/evaluation/)

North Carolina Department of Public Instruction. (2017b). School Performance. Retrieved

March 29, 2017, from <http://www.ncpublicschools.org/src/guide/performance/>

- North Carolina Department of Public Instruction. (n.d.). Teacher performance appraisal-revised fact sheet and summary. Retrieved March 18, 2017 from <http://www.ncpublicschools.org/fbsmanuals/evalpsemployees/EvalFactSheet.pdf>
- North Carolina ESEA Flexibility Request. (2012). Retrieved April 2, 2017, from <https://www2.ed.gov/policy/eseaflex/approved-requests/nc.pdf>
- North Carolina School Report Cards. (2015). Retrieved from [https://ncreportcards.ondemand.sas.com/SASVisualAnalyticsViewer/VisualAnalyticsViewer\\_guest.jsp?reportPath=/ReportCard/NC\\_SRC&reportName=NC+Report+Cards](https://ncreportcards.ondemand.sas.com/SASVisualAnalyticsViewer/VisualAnalyticsViewer_guest.jsp?reportPath=/ReportCard/NC_SRC&reportName=NC+Report+Cards)
- North Carolina School Report Cards. (2018). Retrieved from [https://ncreportcards.ondemand.sas.com/SASVisualAnalyticsViewer/VisualAnalyticsViewer\\_guest.jsp?reportPath=/ReportCard/NC\\_SRC&reportName=NC+Report+Cards](https://ncreportcards.ondemand.sas.com/SASVisualAnalyticsViewer/VisualAnalyticsViewer_guest.jsp?reportPath=/ReportCard/NC_SRC&reportName=NC+Report+Cards)
- North Carolina State Board of Education. (2016). Minutes of the North Carolina State Board of Education. Retrieved March 4, 2017, from <http://stateboard.ncpublicschools.gov/minutes-actions/sbe-minutes/2016-minutes/apr-minutes.pdf/view>
- Neumerski, C.M., Grissom, J.A., Goldring, E., Drake, T.A. (2018). Restructuring instructional leadership: How multiple-measure teacher evaluation systems are redefining the role of the school principal. *The Elementary School Journal*, 119(2), 270-297.
- Olson, L. (1989). In North Carolina, career ladder plan nears a crossroads. *Education Week*. Retrieved March 18, 2017 from <http://www.edweek.org/ew/articles/1989/02/01/08150023.h08.html>
- Parcel, T.L., & Dufur, M.J. (2001). Capital at home and at school: Effects on student achievement. *Social Forces*, 79(3), 881-911.

- Paufler, N.A., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal*, 51(2), 328-362.
- Phillips, K. (2010). What does “highly qualified” mean for student achievement? Evaluating the relationships between teacher quality indicators and at-risk students’ mathematics and reading achievement gains in first grade. *The Elementary School Journal*, 110(4), 464-493.
- Public Schools of North Carolina. (2013). Measuring student learning for educator effectiveness. Retrieved September 4, 2016, from <http://www.dpi.state.nc.us/docs/effectiveness-model/student-growth/measuring-growth-guide.pdf>
- Rice, J.K., Malen, B., Baumann, P., Chen, E., Dougherty, A., Hyde, L., Jackson, C., Jacobson, R., & Mckithen, C. (2012). The persistent problems and confounding challenges of educator incentives: The case of TIF in Prince George's County, Maryland. *Educational Policy*, 26(6), 892-933.
- Rockoff J.E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *The American Economic Review*, 100(2), 261-266. Retrieved from <http://www.jstor.org.prox.lib.ncsu.edu/stable/27805001>.
- Rose, R.A., Henry, G.T., & Lauen, D.L. (2012). *Executive summary of comparing value added models for estimating teacher effectiveness*. Consortium for Educational Research and Evaluation– North Carolina. Retrieved March 29, 2017, from <http://www.ncpublicschools.org/docs/effectiveness-model/evaas/selection/exec-summary-effectiveness.pdf>
- Sanders, W.L. (1998). Value added assessment. *The School Administrator*, 55(11), 24-32.

- 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.
- Sartain, L., Stoelinga, S.R., & Brown, E.R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. University of Chicago Consortium on Chicago School Research.
- SAS Institute, Inc. (2016a). Reports. Retrieved March 4, 2016, from <https://ncdpi.sas.com/welcome.html?as=e>
- SAS Institute, Inc. (2016b). SAS EVAAS for k-12 statistical models. Retrieved March 18, 2017 from [https://www.sas.com/content/dam/SAS/en\\_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf](https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf)
- SAS Institute, Inc. (2017). EVAAS. Retrieved March 18, 2017 from [ncdpi.sas.com](https://ncdpi.sas.com)
- Sass, T.R., Semykina, A., & Harris, D.N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, 38, 9-23.
- Sawchuk, S. (2016). ESSA loosens reigns on teacher evaluations, qualifications. *Education Week*, 35(15), 14-15. Retrieved from <http://www.edweek.org/ew/articles/2016/01/06/essa-loosens-reins-on-teacher-evaluations-qualifications.html>
- Scherrer, J. (2012). What's the value of VAM (value-added modeling)? *Phi Delta Kappan*, 93(8), 58-60.
- Schochet, P.Z., & Chiang, H.S. (2013). What are error rates for classifying teacher and school performance using value-added models? *Journal of Educational and Behavioral Statistics*, 38(2), 142-171.

- Shapiro, E. (2019). New York joins movement to abandon use of student tests in teacher evaluations. *The New York Times*.  
<https://www.nytimes.com/2019/02/01/nyregion/standardized-testing-teachers-students.html?action=click&module=MoreInSection&pgtype=Article&region=Footer&contentCollection=New%20York>)
- Sparks, S.D. (2012). Caution urged in using ‘value-added’ evaluations: Scholars say districts must be more careful. *Education Week*, 32(6), 6.
- Stewart, B.E. (2006). *Value-added modeling: The challenge of measuring educational outcomes*. Carnegie Corporation.
- Taylor, E.S., & Tyler, J.H. (2012). Can teacher evaluation improve teaching? *Education Next*, 12(4), 78-84. Retrieved from  
<http://proxying.lib.ncsu.edu/index.php?url=http://search.proquest.com.prox.lib.ncsu.edu/docview/1237826958?accountid=12725>
- Teacher Performance Appraisal Instrument – Revised (n.d.). Retrieved March 18, 2017 from  
<http://www.ncpublicschools.org/fbsmanuals/evalpsemployees/tpairecu.pdf>
- Templeton, N.R., Willis, K., & Hendricks, L. (2016). The coaching principal: Building teacher capacity through the Texas teacher evaluation and support system (T-TESS). *International Journal of Organizational Innovation (Online)*, 8(4), 140-145.
- Thomas, J.R. (2016). Another push to delay linking teacher evaluations with test scores. *Hartford Courant*. Retrieved April 3, 2017, from <http://www.courant.com/education/hc-teacher-evaluations-test-scores-delayed-20161226-story.html>

- TPAI-R Full Review – Experienced Teachers (n.d.) Retrieved March 18, 2017 from <http://www.ncpublicschools.org/fbsmanuals/evalpsemployees/TPAIR%20Full%20Review.PDF>
- Tracy, S. (1995). How historical concepts of supervision relate to supervisory practices today. *The Clearing House*, 68(5), 320–324.
- Tucker, P.D., & Stronge, J.H. (2005). *Linking teacher evaluation and student learning*. Association for Supervision and Curriculum Development.
- Tufford, L., & Newman, P. (2010). Bracketing in Qualitative Research. *Qualitative Social Work*, 11, 80-96. 10.1177/1473325010368316.
- U.S. Department of Education. (2009). Race to the top program executive summary. Retrieved March 30, 2017, from <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education. (2010). Race to the top application for initial funding: Tennessee. Retrieved March 22, 2017, from <http://www2.ed.gov/prox.lib.ncsu.edu/programs/racetothetop/phase1-applications/tennessee.pdf>
- U.S. Department of Education. (2015). Every student succeeds act (ESSA). Retrieved from <https://www.ed.gov/essa?src=rn>
- Viadero, D. (2008). 'Value added' pioneer says stinging critique of method is off-base. *Education Week*, 27(36), 13.
- Ward, M.E. (1995). Teacher dismissal: The impact of tenure, administrator competence, and other factors. *The School Administrator*, 52(5), 16-19.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our*

- national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project.
- Wetzel, W. (1929). Scientific supervision and curriculum building. *The School Review*, 37(2), 179–192.
- Williams, J.M., McKinney, C., Garland, R., & Goodwin, B. (2010). How North Carolina improved teacher evaluation. *Educational Leadership*, 67(8), 16-23. Retrieved March 21, 2016, from <http://www.ascd.org/publications/educational-leadership/may10/vol67/num08/How-North-Carolina-Improved-Teacher-Evaluation.aspx>
- Wise, E., Darling-Hammond, L., McLaughlin, M., & Bernstein, H. (1984). *Teacher evaluation: A study of effective practices*. RAND.
- Woulfin, S.L., Donaldson, M.L., & Gonzales, R. (2015). District leaders framing of educator evaluation policy. *Educational Administration Quarterly*, 52(1), 110-143.
- Wright, S., White, J., Sanders, W., & Rivers, J. (2010). SAS EVAAS statistical models. Retrieved September 13, 2016 from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.437.6615&rep=rep1&type=pdf>
- Xu, X., Grant, L.W., & Ward, T.J. (2016). Validation of a statewide teacher evaluation system: Relationship between scores from evaluation and student academic progress. *NASSP Bulletin*, 100(4), 203-222.
- Yaffe, D. (2016). Tackling the teacher shortage. *The Education Digest*, 81(8), 11-15.
- Ye Y., Singh K. (2017). The effect of working condition on math teacher effectiveness: value-added scores and student satisfaction in teaching. *Educational Research for Policy and Practice*, 16(3), 283–295. <https://doi.org/10.1007/s10671-016-9207-6>



## APPENDICES

## Appendix A

### Appendix A: Protocol for Principal Interview

#### Introduction

**Say:** *“Thank you for taking the time to speak with me today. I truly appreciate it, and I value your time so I will make sure that I stick to my script and we move through the interview as efficiently as possible.”*

**Say:** *“My name is Phillip Lampron and I am a doctoral student at North Carolina State University. I am conducting a study of principals’ perceptions of both the North Carolina Educator Evaluation System (NCEES) and the Educational Value-Added Assessment System (EVAAS), as well as the correlations between these two systems.”*

**Say:** *“Before we begin, I would like to share a couple of disclosures with you:*

- Your participation in this study is completely voluntary. If at any point you no longer feel comfortable, you may stop the interview at any time.
- You will have an option to either allow me to record the audio of this interview to ensure the validity of your statements, or you may opt to not allow me to record the audio of this interview and I will transcribe the interview as we go. If you choose to allow me to record the audio of this interview the recording will be transcribed, and once it has been transcribed, the recording will be deleted.
- You’ve had a chance to review the consent form and your submission of your name and email address at the end signified your consent to participate. Please respond to the following verbal question:

**Say:** *“Are you still willing to participate in this interview?”*

**If the participant responds “yes,” say:** *“Thank you so much. I will begin the interview.”*

**If the participant responds “no,” say:** *“Thank you for taking time out of your day to speak with me. Enjoy the duration of your day.”*

**Say:** *To facilitate accuracy in note-taking, I would like to audio record our conversations today. For your information, only I, and my committee advisor will have access to the recording, which will be eventually destroyed after it is transcribed. During the recording I will refer to you only by the pseudonym \_\_\_\_\_ in order to further protect your confidentiality.*

*If you prefer that I do not audio record you, I will transcribe the interview by hand and I will refer to you only by the pseudonym \_\_\_\_\_ in order to further protect your confidentiality.*

**Ask:** *May I audio record this interview?*

**For those who give permission to record say:** *To ensure I capture your interview and adhere to validity, I am recording this interview. Recordings will be stored and locked in a safe location. This interview is voluntary and confidential. You will be assigned a pseudonym and I will not disclose your name with your interview responses. I will be taking notes during the interview and you stop this interview at any time without any consequences or penalties. The interview will last approximately thirty minutes. I will be asking you a series of questions relating to your*

*experiences as a principal and your opinions on EVAAS and NCEES correlations. Thank you for agreeing to participate in this study.*

**For those who refuse to record say:** *To honor your audio recording preference, I will not record this interview. To ensure that I capture your interview and adhere to validity, I will be scribing your interview by hand. You can stop this interview at any time without any consequences or penalties. Your scribed interview will be stored and locked in a safe location. This interview is voluntary and confidential. You will be assigned a pseudonym and I will not disclose your name with your interview responses. The interview will last approximately thirty minutes. I will be asking you a series of questions relating to your experiences as a principal and your opinions on EVAAS and NCEES correlations. Thank you for agreeing to participate in this study.*

**For those who give permission to record say:** *The recording will now begin and I will start the interview process. During this process, I will refer to you with a pseudonym to protect your identity.*

**For those who refuse to record say:** *I will start the interview process and scribe your responses. During this process, I will refer to you with a pseudonym to protect your identity.*

**Say:** *“Before I begin asking the interview questions, I would like for us to review the correlations between EVAAS and NCEES of middle school mathematics teachers that have been provided to us by the North Carolina Department of Public Instruction.*

*First, let’s review the state level correlation. (Discuss correlation between EVAAS and NCEES at the state level.)*

*Next, let’s review the district level correlations for your specific school district. (Discuss correlation between EVAAS and NCEES at the district level.)*

*Lastly, I would like for you to review the data that were sent to you by the North Carolina Department of Public Instruction about your specific correlations and reflect on that based on what we have discussed so far. There is no need for you to share your specific correlations with me.”*

### **Background Questions**

1. For how many years have you been a principal in a school in North Carolina?
2. For how many years have you been a principal in your current school district?
3. For how many years have you been a principal at your current school?
4. Approximately what percent of the students in your school receive free or reduced priced lunches?
5. Would you describe the school at which you served during the 2017-2018 school year as urban, suburban, or rural?

### **Formal Interview Questions**

1. Describe the utility and quality of the professional development or training you've had related to NCEES and EVAAS.
2. How well do you believe these two systems align with regards to rating the effectiveness of a teacher?
3. Describe your perspectives regarding the degree to which scores from these systems should be correlated.
4. Based on your review of your NCEES and EVAAS correlations that were sent to you by the North Carolina Department of Public Instruction and your review of the state and district level results that we have reviewed today, how would you describe the correlations at the state level? How about the district level?
5. To what do you attribute the degree of correlation that exists between each level (state, district, and school/principal)?
6. Describe the specific teacher attributes that you feel contribute directly to an increase in student achievement.
7. Describe your perspectives on whether these attributes are adequately addressed in NCEES.
8. Describe your perspectives on whether evaluation of such attributes would affect the correlations among NCEES and EVAAS ratings.

**Appendix B**

**Appendix B: Principal Perceptions of NCEES and Principal Predictions of Teachers' EVAAS Ratings Validity Questionnaire**

Thank you for taking the time to review this instrument. I appreciate your willingness to give your expertise and assistance in the development of this instrument that will be used to gather data for this study. Your input is very important and will be used to make any necessary changes in order to more effectively meet the goal of the study.

The purpose of this instrument you are evaluating is to gather data from elementary principals regarding whether they believe teacher ratings in the North Carolina Educator Evaluation System (NCEES) are accurate, their ability to predict where elementary mathematics teachers will be rated according to the Education Value-Added Assessment System (EVAAS), and to provide input regarding what information they used to make these predictions. It is my desire that this survey will provide insight into whether principals are more accurately able to predict a teacher's impact on student growth than they are willing or able to articulate in NCEES.

Please take your time and critique the attached questionnaire by answering either "Yes" or "No" to the questions below. Please provide feedback as well as your reasoning(s) behind any responses that receive a "No" on the lines provided.

1. Has the survey been developed with language that is easy to understand by the participants in this study?

Yes \_\_\_\_\_ No \_\_\_\_\_

---

---

---

---

2. Does the survey address suitable issues that will allow the researcher to obtain pertinent information related to principal perceptions of NCEES and principal predictions of teachers' EVAAS ratings?

Yes \_\_\_\_\_ No \_\_\_\_\_

---

---

---

3. Are there any particular items within the survey that you would modify?

Yes \_\_\_\_\_ No \_\_\_\_\_

(Please identify item number(s) if you selected a “Yes” response.)

---

---

---

---

4. Do you believe any of the survey items can be potentially offensive or invasive to the participant?

Yes \_\_\_\_\_ No \_\_\_\_\_

(Please identify item number(s) if you selected a “Yes” response.)

---

---

---

---

5. Are there any items in the survey that you feel should be excluded?

Yes \_\_\_\_\_ No \_\_\_\_\_

(Please identify item number(s) if you selected a “Yes” response.)

---

---

---

---

6. Are there any survey items that you feel should be included that are **not** included in the attached questionnaire?

Yes \_\_\_\_\_ No \_\_\_\_\_

(If you selected "Yes", please write your suggested statement(s) below.)

---

---

---

---

---

7. Please feel free to provide any further suggestions or comments that you feel would strengthen the validity of this survey on the lines below.

---

---

---

---

---

---

## Appendix C

### Appendix C: LETTER TO THE SUPERINTENDENT

Date  
Name of Superintendent  
Name of School District  
Address

RE: Permission to Conduct Research Study

Dear Superintendent \_\_\_\_\_,

My name is Phillip Lampron, and I am currently enrolled in the Educational Administration and Supervision doctoral program at North Carolina State University. In order to complete my dissertation for this program, I need to conduct surveys of principals of middle schools in the Northeast and North Central Regions of North Carolina. The goal of my dissertation is to determine to what extent correlations exist between ratings of teachers in the North Carolina Educator Evaluation System (NCEES), the Education Value-Added Assessment System (EVAAS), and principal predictions of teachers' EVAAS ratings. Principals will also be asked to provide some background about the school in which they serve, their leadership practices, and their beliefs about principal practices in general. In order for these correlations to be ascertained, principals will be asked to provide the unique identification numbers (UIDs) for middle school mathematics teachers whom they were responsible for evaluating in the NCEES system during the previous school year. The UIDs of teachers will only be used to conduct data analysis, and will be erased once those analyses are performed. Neither the UIDs nor names of teachers will be used for reporting purposes of this study. If you agree to participate, I will also be sending you a Memorandum of Agreement from the North Carolina Department of Public Instruction that would allow them to release to me individual teacher ratings from both NCEES and EVAAS for your district. The data that are released to me will not contain teacher names, but solely their UID number and their ratings in these two systems. This will allow me to compare principal predictions with their actual EVAAS results when they are released. This study has been approved by the Institutional Review Board of North Carolina State University.

The purpose of this letter is to request permission to survey middle school principals in your school district. With your approval, I will email the middle school principals in your district a link that will take them to the survey that has been created for this dissertation study. All survey responses will remain anonymous at all times. No teachers, principals, schools or districts will be identified anywhere in the research findings.

Please feel free to contact me if you have any questions or concerns at (919) 738-1507 or palampro@ncsu.edu. My dissertation co-chairs are Dr. Mike Ward and Dr. Lance Fusarelli. They can be contacted at meward@ncsu.edu and ldfusare@ncsu.edu, respectively.



If you agree to my request, please sign and return the form on the second page of this document. The signed form can be emailed back to me at [palampro@ncsu.edu](mailto:palampro@ncsu.edu), or faxed to (252) 641-1816.

Sincerely,

Phillip A. Lampron

Consent Form

By signing and returning this form, I give Phillip Lampron, a doctoral candidate at North Carolina State University, permission to conduct a research study in the \_\_\_\_\_ District. I acknowledge that Phillip Lampron may contact the middle school principal(s) via email to solicit survey responses during the months of July 2018 – September 2018. I also understand that a Memorandum of Agreement will be forthcoming with the Department of Public Instruction that will allow Phillip Lampron to be providing with the ratings of teachers in my district from both the North Carolina Educator Evaluation System (NCEES) and the Education Value Added Assessment System (EVAAS).

Approved by:

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

Superintendent Name  
District  
Address

## Appendix D

### Appendix D: Principal Consent Form

**Title of Study:** An Examination of Consistency between Objective and Subjective Measures of Teacher Performance (eIRB #: 20646)

**Principal Investigator(s):** Phillip Lampron, palampro@ncsu.edu, (919) 738-1507

**Funding Source:** None

**Faculty Point of Contact:** Lance D. Fusarelli, [ldfusare@ncsu.edu](mailto:ldfusare@ncsu.edu) or (919) 513-0507; Michael E. Ward, [meward@ncsu.edu](mailto:meward@ncsu.edu), (919) 710-7317

---

#### **What are some general things you should know about research studies?**

You are invited to take part in a research study. Your participation in this study is voluntary. You have the right to be a part of this study, to choose not to participate, and to stop participating at any time without penalty. The purpose of this research study is to gain a better understanding of the perceptions of principals about correlations between the Education Value-Added Assessment System (EVAAS) and the North Carolina Educator Evaluation System (NCEES). We will do this through reviewing data provided by the North Carolina Department of Instruction and interviews with principals.

You are not guaranteed any personal benefits from being in this study. Research studies also may pose risks to those who participate. You may want to participate in this research because you will be adding to the body of knowledge about teacher performance metrics in the field of education. You may not want to participate in this research because you might be re-identified due to the information you share.

Specific details about the research in which you are invited to participate are contained below. If you do not understand something in this form, please ask the researcher for clarification or more information. A copy of this consent form will be provided to you. If, at any time, you have questions about your participation in this research, do not hesitate to contact the researcher(s) named above or the NC State IRB office. The IRB office's contact information is listed in the *What if you have questions about your rights as a research participant?* section of this form.

#### **What is the purpose of this study?**

The purpose of the study is to review principal perceptions in the Northeast and North Central regions of North Carolina about correlations between EVAAS and NCEES ratings.

#### **Am I eligible to be a participant in this study?**

There will be approximately seven participants in this study.

In order to be a participant in this study, you must agree to be in the study and be a principal for a middle school in the Northeast or North Central regions of North Carolina.

You cannot participate in this study if you do not want to be in the study or if you are no longer a principal of a middle school in the Northeast or North Central region of North Carolina.

### **What will happen if you take part in the study?**

If you agree to participate in this study, you will be asked to do all of the following:

1. Schedule a time to participate in an interview with me in a private location via phone or online platform during the COVID-19 pandemic. This location will be outside of the workplace.
2. Review correlations between EVAAS and NCEES ratings of your district, which will be provided for you.
3. Participate in an interview with the principal investigator which will last approximately 30 minutes.

The total amount of time that you will be participating in this study is approximately 35 minutes.

### **Recording**

As a part of this research, I would like your consent to audio record you. You may choose to not be recorded; in that event, I will take extensive notes during the interview. Please provide your response to the question of being audio recorded in the appropriate space at the end of this document.

### **Risks and benefits**

There are minimal risks associated with participation in this research. The risks to you as a result of this research are related to your employment and the potential of your identity and responses becoming known to your employer. I am mitigating this risk as much as possible by removing direct identifiers from the data prior to publication, reporting interview data with all identifiers stripped, and reporting quotes from the interview in such a manner that your name and the name of your school or district will not be on the data I publish.

There are no direct benefits to your participation in the research. The indirect benefits are that you will be adding to the body of knowledge in the field of education regarding objective and subjective measures of teacher evaluation.

### **Right to withdraw your participation**

You can stop participating in this study at any time for any reason. In order to stop your participation, please contact Phillip Lampron at [palampro@ncsu.edu](mailto:palampro@ncsu.edu) or (919) 738-1507, or Lance D. Fusarelli at [ldfusare@ncsu.edu](mailto:ldfusare@ncsu.edu) or (919) 513-0507, or Dr. Mike Ward at [meward@ncsu.edu](mailto:meward@ncsu.edu) or 919-710-7317. If you choose to withdraw your consent and to stop participating in this research, you can expect that the researcher(s) will redact your de-identified information from their data set, securely destroy your data, and prevent future uses of your de-identified information for research purposes wherever possible. This is possible in some, but not all, cases.

## **Confidentiality, personal privacy, and data management**

Trust is the foundation of the participant/researcher relationship. Much of that principle of trust is tied to keeping your information private and in the manner that we have described to you in this form. The information that you share with me will be held in confidence to the fullest extent allowed by law.

Protecting your privacy as related to this research is of utmost importance to me. However, there are very rare circumstances related to confidentiality where I may have to share information about you. These are limited to instances in which imminent harm could come to you or others.

How we manage, protect, and share your data are the principal ways that I protect your personal privacy. Data that will be shared with others about you will be re-identifiable.

**Re-identifiable.** Re-identifiable data is information that I can use to identify you indirectly because of my access to information, role, skills, combination of information, and/or use of technology. This may also mean that in published reports others could identify you from what is reported, for example, if a story you tell us and that I relate in the reporting of findings is very specific. If your data is re-identifiable, I will report it in such a way that you are not directly identified in reports. Based on how we need to share the data, I cannot remove details from the report that would protect your identity from ever being figured out. This means that others may be able to re-identify from the information reported from this research.

To help maximize the benefits of your participation in this project, destruction of master lists of names, emails, audio recordings, and transcripts will take place at the end of the study. Your re-identifiable information will not be kept for future research purposes by this or other researchers.

### **Compensation**

For your participation in this study, you will receive no compensation

### **What if you have questions about this study?**

If you have questions at any time about the study itself or the procedures implemented in this study, you may contact the researcher, Phillip Lampron, at [palampro@ncsu.edu](mailto:palampro@ncsu.edu) or (919) 738-1507. You may also contact the faculty advisors, or Dr. Mike Ward at [meward@ncsu.edu](mailto:meward@ncsu.edu) or 919-710-7317 or Lance D. Fusarelli, at [ldfusare@ncsu.edu](mailto:ldfusare@ncsu.edu) or (919) 513-0507.

### **What if you have questions about your rights as a research participant?**

If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact the NC State IRB (Institutional Review Board) Office. An IRB office helps participants if they have any issues regarding research activities. You can contact the NC State IRB Office via email at [irb-director@ncsu.edu](mailto:irb-director@ncsu.edu) or via phone at (919) 515-8754.

## **Consent To Participate**

By inserting my name, email address, audio recording preference, and available dates for an interview and returning this information to Phillip Lampron's email address, I am affirming that I have read and understand the above information. All of the questions that I had about this research have been answered. I have chosen to participate in this study with the understanding that I may stop participating at any time without penalty or loss of benefits to which I am otherwise entitled. I am aware that I may revoke my consent at any time.

### **Participant Information:**

Participant's name: \_\_\_\_\_

Participant's email: \_\_\_\_\_

### **Check the appropriate space below:**

I consent to being audio recorded during the study interview: \_\_\_\_\_

I do not consent to being audio recorded during the study interview: \_\_\_\_\_

### **Availability for interview:**

I am available for an interview on the following dates and times:

\_\_\_\_\_

