Reading and Vocabulary Value-Added Scores and Literacy Teaching Quality
Investigating the Relationship Between 2nd Grade Teachers' Value-Added Scores and Teaching Quality

Melissa Jane Luck
READING AND VOCABULARY VALUE-ADDED SCORES AND LITERACY TEACHING QUALITY

INVESTIGATING THE RELATIONSHIP BETWEEN 2ND GRADE TEACHERS’ VALUE-ADDED SCORES AND TEACHING QUALITY

By

MELISSA JANE LUCK

A dissertation submitted to the School of Teacher Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Degree Awarded:
Summer Semester, 2012
Melissa Luck defended this dissertation on June 14th, 2012.

The members of the supervisory committee were:

Carol Connor
Professor Directing Dissertation

Christopher Schatschneider
University Representative

Sherry Southerland
Committee Member

Young Suk-Kim
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.
For Sebastian

Everything I do is for you. I hope that you never stop

learning, always love reading

and know that anything is possible.
ACKNOWLEDGEMENTS

I would like my family to know how much I have appreciated their support and love during this journey. I could not have done it without the encouragement of my husband, David, my mother, Theresa, and my son, Sebastian.

Dr. Connor, your continued support and academic guidance enabled my success. The writing of my dissertation would not have happened without your help and remarkable mentoring. I have learned so much from you. Thank you for everything.

Dr. Schatschneider, I began my knowledge of statistics in your introductory class several years ago and am thankful to have received such excellent instruction. Thank you for your suggestions to improve my methods and your help throughout.

Dr. Southerland, I thank you for your time and willingness to be a part of my committee. I appreciate your valuable insight in the design of my study.

Dr. Kim, I learned how to administer reading assessments in your class, in addition to the nuances required for score interpretation and establishing interventions. Thank you for the academic support you’ve provided me.

Dr. Al’Otaiba, although not a member on my final committee, provided support by participating in both my preliminary and prospectus defense committees. I am thankful for her guidance as well.

A special thanks to all of the members of the iSi lab, teachers and students who participated in this study. Dr. Phyllis Underwood, Jennifer Dombek, Novell Tani, Stephanie Glasney, Lauren White, Meghann Montgomery and others all played a role in helping this study to become a reality. Additionally, a much-deserved thanks goes to Allison Ames for all of her dedication to this project. Thank you for your hard work and dedication to conducting exceptional research.

Last, but certainly not least, I would to thank the women who provided childcare to my baby, Osiris Alaniz and Delia Maldonado. Particularly Delia, you are an angel. Without uninterrupted time to work (that you so kindly provided me), this dream of writing my dissertation would have never become a reality. I am indebted to you.
TABLE OF CONTENTS

List of Tables .................................................................................................................. viii
Abstract ............................................................................................................................. x

1. CHAPTER ONE ........................................................................................................... 1
   1.1 Value-Added Methods ......................................................................................... 4
   1.2 Teacher Quality and Student Achievement ...................................................... 5
   1.3 Teacher Quality and Educational Legislation ................................................... 8
   1.4 Defining Teaching Quality .................................................................................. 10
   1.5 Measuring Teaching Quality .............................................................................. 11

2. CHAPTER TWO ......................................................................................................... 12
   2.1 Value-Added Methods ......................................................................................... 13
   2.2 Benefits of Value-Added Methods ....................................................................... 13
   2.3 Limitations of Value-Added Methods .................................................................. 14
   2.4 Modeling Value-Added ....................................................................................... 17
   2.5 What Contributes to Teacher’s Value-Added Score? ........................................... 21
   2.6 Teacher-Level Factors ....................................................................................... 21
   2.7 Teaching Experience ......................................................................................... 21
   2.8 Teacher Knowledge ............................................................................................ 23
   2.9 Quality of Teaching ............................................................................................ 23
   2.10 Teacher Perception of Student Social Skills .................................................... 24
   2.11 School-Level Factors ....................................................................................... 25
   2.12 Student Level Factors ...................................................................................... 26
   2.13 Classroom Composition .................................................................................... 26
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.13</td>
<td>Difference from Modal Age</td>
<td>28</td>
</tr>
<tr>
<td>3.</td>
<td>CHAPTER THREE</td>
<td>29</td>
</tr>
<tr>
<td>3.1</td>
<td>Purpose</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>Research Questions</td>
<td>29</td>
</tr>
<tr>
<td>3.3</td>
<td>Study Design</td>
<td>30</td>
</tr>
<tr>
<td>3.4</td>
<td>Participants and Setting</td>
<td>30</td>
</tr>
<tr>
<td>3.5</td>
<td>Data Collection Procedures</td>
<td>31</td>
</tr>
<tr>
<td>3.6</td>
<td>Teacher-Level Measures</td>
<td>31</td>
</tr>
<tr>
<td>3.7</td>
<td>Teaching Experience</td>
<td>31</td>
</tr>
<tr>
<td>3.8</td>
<td>Teacher Knowledge</td>
<td>31</td>
</tr>
<tr>
<td>3.9</td>
<td>Teacher’s Perception of Students’ Skills</td>
<td>32</td>
</tr>
<tr>
<td>3.10</td>
<td>Quality of Teaching</td>
<td>32</td>
</tr>
<tr>
<td>3.11</td>
<td>Classroom-Level Measures</td>
<td>33</td>
</tr>
<tr>
<td>3.12</td>
<td>Student Self-Regulation</td>
<td>33</td>
</tr>
<tr>
<td>3.13</td>
<td>Student-Level Measures</td>
<td>34</td>
</tr>
<tr>
<td>3.14</td>
<td>Reading and Vocabulary Skills</td>
<td>34</td>
</tr>
<tr>
<td>3.15</td>
<td>Data Analysis Procedures</td>
<td>35</td>
</tr>
<tr>
<td>3.16</td>
<td>Multi-Level Models</td>
<td>35</td>
</tr>
<tr>
<td>3.17</td>
<td>Analytic Strategies</td>
<td>38</td>
</tr>
<tr>
<td>4.</td>
<td>CHAPTER FOUR</td>
<td>41</td>
</tr>
<tr>
<td>4.1</td>
<td>Results</td>
<td>41</td>
</tr>
<tr>
<td>5.</td>
<td>CHAPTER FIVE</td>
<td>54</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary of Findings</td>
<td>54</td>
</tr>
</tbody>
</table>
5.2 Limitations........................................................................................................57
5.3 Conclusion.........................................................................................................58

6. CHAPTER SIX......................................................................................................61
   6.1 Appendix A: Teacher Knowledge Survey..................................................61
   6.2 Appendix B: Teacher Background Questionnaire ......................................69
   6.3 Appendix C: Teacher Fidelity Rubric ..........................................................74
   6.4 Appendix D: Human Subject Consent Forms ............................................81
   6.5 Appendix E: IRB Approval Letter ...............................................................89
   6.6 Bibliography ...............................................................................................91
   6.7 Biographical Sketch ....................................................................................100
LIST OF TABLES

1.1 Models Employed and Summary of Their Characteristics 55
1.2 Mean VAS by Model and Quintile 59
1.3 Percentage of Agreement for SFEM PC VAS, HLM PC, and CAM PC VAS 60
1.4 Percentage of Agreement for SFEM PV VAS, HLM PC, and CAM PV VAS 60
1.5 Percentage of Agreement for HLM PC/CAM PC and HLM PV/CAM PV VAS 62
1.6 Percentage of Agreement between SFEM PC and SFEM PV 63
1.7 ICC Analysis 56
2.1 Descriptive Statistics 64
2.2 Correlation Coefficients 65
2.3 Regression Results 66
2.4 Multiple Correlation Analysis Results for PC Models 69
2.5 Multiple Correlation Analysis Results for PV Models 70
ABSTRACT

Teacher quality has emerged as a key factor influencing the success of students, both academically, and beyond. The national discussion of teacher quality focuses on determining which teachers are effective and how best to determine which qualities or characteristics lead to increased student achievement gains. In 2012, value-added methods (VAM) have become education’s answer to the question of determining teacher efficacy.

Value-added methods seek to isolate the contribution teachers have on students’ learning gains, which can be compared to the performance measures of other teachers. However, there are many different, competing methods from which to choose. The problem lies with the fact that a consensus has yet to be reached regarding an acceptable method for calculating these value-added scores (VAS). This study investigated the validity of three different VAM by comparing their resulting VAS with other established measures of teaching quality.

Results revealed that the three value-added methods did not consistently predict teaching quality (TQ) across all models. Specifically, SES and teacher knowledge were identified as significantly positive predictors for one of the three models (SFEM) when regression analyses were run. Interestingly, the variables typically expected to be associated with higher VAS (teacher knowledge TK and classroom environment TQ), were not associated with higher VAS for Passage Comprehension (PC).

The results of this study highlight clear concerns with VAM, albeit with a small sample of teachers (n=7). At the same time, the nation’s reliance on VAM continues to increase. Recommendations include a universal, standardized method for calculating VAS.
CHAPTER ONE
INTRODUCTION

Reading and Vocabulary Value-Added Scores and Literacy Teaching Quality
Investigating the relationship between 2nd Grade Teachers’ Value-Added Scores and Teaching Quality

The current national discussion of teacher quality is perhaps one of the most important and recurring issues in education (Haskins & Loeb, 2007). This discussion of teacher quality primarily focuses on how best to determine which teachers are effective and which qualities or characteristics lead to increased student achievement gains, often considered the ultimate measure of effective instruction (Goe, 2007). Within the context of accountability measures and value-added methods, student academic achievement is defined as outcomes on standardized achievement tests measuring specific academic skills areas, such as reading (Malecki & Elliott, 2002). For the purpose of educational accountability, gains in student achievement are typically indicators of teacher quality. Value-added methods (designed to measure teacher effects) produce a score for teachers that rely on these gains in student learning. Under this definition of teacher quality, student learning equals an increase in standardized test-scores. The logic behind this idea is that high quality teachers should be able to help students learn more effectively, as evidenced by student gain scores from year to year (Huang & Moon, 2009). Value-added methods seek to isolate the contribution individual teachers have on their students’ learning gains. Thus, the purpose of this study is to examine literacy teacher value-added scores using a variety of methods, examine how consistently they identify the same teachers, and whether they agree with well-regarded measures of instructional practice based on classroom observation.

In the past, certain school and teacher performances have been assessed based only on students’ standardized test scores from the current year (i.e., status score) to estimate school or teacher effects on student performance. These methods, known as
status-based methods typically rely on regression models, and may or may not include teacher, student or school-level variables that could influence test scores. The distinguishing characteristic of status-based methods is the absence of adjustment for students’ incoming knowledge level. The obvious deficiency of such methods is that differences among schools in average knowledge of incoming students would confound the assessment of instructional quality at each school, and, consequently, the quality of teaching in individual classrooms. Because of this flaw in status-based methods, alternatives that adjust for incoming differences in knowledge, such as value-added measures are generally preferred (Tekwe et al., 2004).

Currently, within the educational policy community, there is a growing trend to use changes in student test-scores to evaluate teaching efficacy. This approach, known as value-added methods (VAM) is gaining popularity among policy-makers as a potentially more objective way of satisfying accountability demands. VAM seek to isolate the contribution individual teachers have on students’ learning gains, which can be compared to the performance measures of other teachers. However, a consensus regarding the most suitable model for calculating value-added scores (VAS) to determine teacher quality has yet to be reached. This is particularly important, as there are numerous methods available for determining VAS. Although VAM are currently a popular method for evaluating teaching quality, to date, a theory of action linking VAM to improved teaching quality does not exist. Therefore, this study seeks to determine the validity of three different value-added methods by comparing their results with other established measures of teaching quality including teacher knowledge (TK) and classroom observations.

Presently there are two popular types of methods for calculating VAS for teachers; gain score models and residual score models (regression models). An example of a commonly used gain score model is the Simple Fixed Effects Model (SFEM, Tekwe, et al., 2004). Two widely used residual score models are the Hierarchical Linear Model Method (HLMM, Tekwe et al., 2004) and the Covariate Adjustment Model (CAM, AIR, 2011). Further descriptions of each model are discussed in the literature review chapter and statistical details are provided in the methods chapter.

Factors that may contribute to higher or lower quality teaching include a variety of teacher, student, and school-level variables. Teacher-level variables may include
content-area knowledge, years of experience, and educational attainment. Student-level variables may include academic proficiency, student characteristics and classroom composition. School-level variables include factors such as Socioeconomic Status (SES), among others. These variables combine to create a unique learning environment for students, and a distinct instructional environment for teachers. It is logical to assume that each of these different variables contribute to teacher’s value-added scores. Hence, knowing which factors influence teacher’s VAS has important implications for both policy and practice, because future efforts to ensure high-quality teaching in every classroom may rely on these identified factors. This study seeks to understand the extent of influence each identified factor might have on a teacher’s VAS, in addition to ascertaining which model(s) identify teachers considered high quality by established measures of teaching quality.

Emerging evidence suggests that teacher effects persist by accumulating over a child’s school career (Sanders & River, 1996; Mendro et al., 1998; Rivers, 1999), meaning that the benefit of good (high-quality) teachers, as well as the detriment of bad (poor quality) teachers follows students throughout their academic careers. This evidence underscores the need to accurately identify those teachers who are effective, as well as those who are less effective and to provide training to the latter group to ensure strong student outcomes. This study seeks to identify the value-added method(s) that appear to be valid measures for determining teaching quality (TQ). As Taylor and colleagues noted, putting high-quality teachers in the classroom will not eliminate variability among students nor guarantee equally high achievement from all children, but ignoring teachers as salient contributors to the classroom environment represents a missed opportunity to promote children’s potential in school and their success in life (Taylor, et al., 2010). Additionally, recent research has found that non-random assignment of students to teachers can bias value added estimates of teachers’ causal effects (Rothstein et. al., 2009). Hence, value-added methods attempt to determine TQ by examining student achievement in relation to other factors but do have some limitations.

Whereas most education researchers and policy analysts agree regarding the primary role that teachers play in advancing student achievement, they are often at odds over the best means to identify effective teachers and improve teaching quality (Berry,
Currently, teachers are being judged by their students’ test-scores via value-added methods as one but increasingly popular method of evaluating teacher quality and satisfying accountability demands. Another potential concern with value-added scores is that student outcomes used to compute the scores are too far removed from the instruction that is being provided. For example, elementary teachers might be assessed on their students’ reading and mathematics scores using a general achievement tests. Additionally, there is far less agreement with regard to accepted standardized assessments for teachers in grades beyond the elementary level, particularly in subjects such as social studies and science. Moreover, it remains to be determined how well these different value-added methods agree with one another with regard to their ability to identify quality teaching. This study will evaluate three value-added methods using student outcomes in reading and vocabulary that are closely aligned with the content and skills being taught during the dedicated block of time devoted to literacy; and examine which model(s) are likely to predict other, established measures of teaching quality, including classroom observations, in an effort to compare the strengths and weaknesses of each model. In the following sections, key issues will be highlighted in the areas of (1) value-added methods; (2) teaching quality and student achievement; (3) teaching quality and educational legislation, and (4) defining and measuring teacher quality.

**Value-Added Methods**

The question of how to estimate teacher contributions to student learning is fundamental to both educational policy and practice. At a time when teacher quality has clearly emerged as a key factor in student learning, a statistical technique that determines the *added value* that teachers bring to student achievement is now under review (Rotham, 2010). Value-added measures can compare students’ growth in achievement to their expected growth, based on prior achievement and demographic factors while controlling for non-school influences such as race, gender, or student characteristics, among others. Under the value-added approach, an expected rate of growth for each student (academic trajectories) may be statistically calculated based on the students’ achievement and other factors. Each student’s actual achievement is then compared to the expected rate of achievement to determine the *value added* by a particular teacher (Meyer, 1997).
Generally, teachers whose students exceed the expected rate of growth are considered highly effective, whereas those whose students grow in achievement at a slower rate are considered less effective. Some advocates of the value-added approach have proposed it as a fair way of awarding teachers monetary incentives such as bonuses based on student performance, and some school districts are using value-added methods in pay-for-performance systems. Other districts have found the assessments offer a great deal of information that can lead to school improvement, and they are using the method for tracking school performance, determining professional development needs and identifying effective teachers without directly linking teacher pay to value-added scores (Rotham, 2010).

**Teaching Quality and Student Achievement**

Central to the study of teacher quality and teacher effects is the idea that the amount students learn in a given academic year is partially a result of their teachers (Huang & Moon, 2009). The available research on teacher effects is particularly useful in aiding our understanding of value-added methods, as VAM seek to identify these teacher effects. Several studies have found the impact of good teachers as evidenced by student achievement (Sanders & Rivers, 1996; Darling-Hammond, 2000). According to these data, teacher quality is the single most important feature of a school that drives student achievement (Haskins & Loeb, 2007). Therefore, the quality, or lack thereof of classroom teaching has a direct impact on student learning, and consequently, student achievement.

Hanushek (1992) estimated that the difference between a student having a *good* teacher and having a *bad* teacher can be more than one grade-level equivalent in test score performance. As a result, students who receive instruction with *good or bad* teachers for several years in a row can have very different learning growth trajectories, after only a few years of schooling. Emerging evidence suggesting that the effects of good instruction accumulate over a child’s school career further underscores the importance of teachers (Sanders & Rivers, 1996; Mendo et al., 1998; Rivers, 1999). However, one problem lies with how educators, researchers and policymakers define *good* and *bad*. The findings from Hanushek (2002) indicate how vital it is to the health of our education system to discern which characteristics yield high quality teaching.
More recently, Hanushek (2005) estimated that the value of having a good teacher (those placed in the 85th percentile or above in terms of quality) might overcome the average achievement deficit between low-income students and others (thus, potentially narrowing the ever-widening achievement gap). This evidence is quite staggering and highlights the importance of understanding which characteristics contribute to quality teaching.

Encouragingly, there is also research to suggest that the opposite is true for student gains. Gordon et al. (2006) found that students taught by teachers in the top quartile of effectiveness advance, on average, approximately five percentile points each year relative to their peers, whereas those taught by teachers in the bottom quartile of effectiveness lose, on average, five percentile points relative to their peers. Imagine the cumulative impact on student learning when children consecutively have either effective or ineffective teachers. This is an alarming reality for our children, given that these teacher effects appear to persist well into the future (McCaffrey et al., 2004). For these reasons, an investigation into the validity of several VAM is essential to the quality of our educational system for the purposes of evaluating teaching quality.

Additionally, Sanders and Horn (1998), found that teacher effectiveness was the single most important factor influencing student learning, and consequently, the academic growth of a student. In their replicated study design, they modeled gains in student test scores as a function of random teacher effects and a small set of student covariates, including achievement. They standardized the contributions of all variables in the models using z-scores. Then, through meta-analysis, the authors reviewed the results of the 30 replicated models and found that the z-scores for teacher effects exceeded the standardized contribution of every other variable in 26 of 30 models. The authors conclude that teachers do, in fact affect student learning, confirming the extent to which teachers impact student knowledge.

Similarly, Rivkin et al. (2000) conducted a study utilizing multiple cohorts of students, each with three consecutive years of test scores, in an effort to remove the effects of student achievement while isolating the impact of teachers on student learning. The authors concluded that teacher effects do exist and estimate, that, (as a lower bound), teachers account for 3.2 percent of variance in student achievement. Alternatively, this variance equals a one-standard-deviation increase in teacher effectiveness is associated
with about a 0.18-standard-deviation increase in student test scores. This evidence suggests the powerful impact effective teachers have with their students and why it is so imperative to identify these teachers.

In another study (Rowan et al., 2002), the authors found that residual classroom-level variance accounts for a significant proportion of variability in student achievement score growth. The authors found robust results across subjects (reading and math), statistical models, and two cohorts of students from a nationwide sample of schools. Although classrooms accounted for meaningful portions of the variance in all models, the magnitude of the variance explained differed. The results of this study also highlight the importance of teacher effects on student achievement (as evidenced by increased standardized test scores).

Rivkin et al. (2005) found that teachers made a difference in student test performance and that student achievement gains were systematically related to teacher characteristics. Likewise, Sanders (1998) concluded, the single largest factor affecting student academic growth is differences in effectiveness of individual classroom teachers. Sander’s conclusion is based on the finding that differential teacher effectiveness is a strong determinant of discrepancies in student learning (Darling-Hammond, 2000). These studies highlight the fact that teachers can have a powerful influence on student learning, and emphasize the importance of establishing a standardized value-added method for calculating teaching quality. Although researchers generally agree that teachers make a difference in student learning, education research has failed to reach a consensus as to what readily identifiable teacher characteristics enable educators to foster increased student achievement gains (Goldhaber & Anthony, 2007; Palardy & Rumberger, 2008). Thus, this study seeks to determine the merits of several methods of calculating reading and vocabulary value-added scores for the purposes of evaluating teaching quality in the domain of literacy education.
Teacher Quality and Educational Legislation

The question of how to quantify effective teaching continues to be heavily debated within the educational and political communities. In 2002, the United States government passed into law the *No Child Left Behind* Act (NCLB), a piece of legislation that provided the federal government with a mandate for education reform on an unprecedented scale (US Department of Education, 2002a). It links government funding to strict improvement and accountability policies for America’s public schools (Smith & Gorard, 2007). For the nation’s public school teachers, NCLB means complex systems of performance and accountability measures aimed at addressing concerns over teacher quality and increasing the number of highly qualified teachers in America’s schools (2007). Specifically, NCLB requires public reporting of teacher qualifications, placing teacher’s characteristics at the forefront of the national discussion on teacher quality.

As it applies expressly to teacher quality, NCLB has two key objectives. The first is to ensure that every teacher is highly qualified in the subject(s) they teach (Smith & Gorard, 2007). The second is to reduce the barriers to becoming a teacher by ‘retooling’ traditional teacher education programs and creating non-traditional routes into the profession (i.e., alternative certification and programs such as Troops to Teachers).

NCLB expressly identifies teacher qualifications as an important component leading to stronger student achievement outcomes. With NCLB’s push for highly qualified teachers, the question of what constitutes quality teaching becomes even more imperative (Huang & Moon, 2009). Teacher quality is also a key provision of the 2004 NCLB renewal legislation, which requires states to provide students with a *highly qualified* teacher in all core courses (core courses include reading, mathematics and science) in every classroom. It is important to note that under the NCLB Act, *highly qualified* teachers are defined as possessing both subject matter (content area) and pedagogical knowledge (Berry et. al, 2004). In accordance with NCLB, teachers must demonstrate that they have sufficient content knowledge for the subject(s) they teach, as evidenced by satisfactory performance on state-level certification examination(s). The legislation stipulates that *highly qualified* teachers must obtain full state certification (including certification obtained through alternative routes) by passing the State teacher licensing examination, and hold a license to teach in the State in which they seek
employment (US Department of Education, 2002b, p. 4). Note that in most states, simply obtaining a passing score on a certification examination is necessary but not sufficient to acquire full teaching certification. In Florida, for example, it is first necessary to pass the General Knowledge (GK) test, then teacher candidates are required to pass a subject-specific test (content-area), such as the Elementary Education Examination (K-6). The K-6 exam assesses content-area knowledge in a variety of subjects (ranging from core subjects such as reading and math, and science and history, as well as the humanities, including music, art, and physical education), and enables candidates to apply for their teaching certification. Research suggests that the quality of instruction students receive directly impacts their learning and, consequently, their ability to succeed academically. Thus, the consensus is clear: high-quality instruction is needed for students to effectively learn, hence the current requirements for teacher licensing.

Most recently, Florida governor Rick Scott signed the Florida Merit Pay Law on March 24th, 2011 (State of Florida, Department of State, 2011, SB 736). Florida is a pioneer in the effort to base teacher salaries on student performance (by providing the teachers with the best results the greatest monetary incentives), as this is one of the first states to adopt such legislation. Merit pay is a reform pillar of former Florida Governor Jeb Bush and the Foundation for Excellence in Education he founded. The idea is to reward teachers who receive the best results (as evidenced by improvement in standardized achievement scores) from their students by offering permanent salary increases, just as the private sector would pay for performance.

Knowing which VAM are valid measures of determining teaching quality is an essential component in providing our students with a high quality education. If students do not receive quality instruction, then educators and researchers alike will not be able to determine a child’s ability to learn. High quality teachers and instruction are the cornerstones of ensuring that children are given adequate opportunities to succeed academically. Ultimately, effective instruction will allow each student to reach his or her maximum potential (Kupermintz, 2003).
Defining Teaching Quality

One difficulty educators encounter is that the term *teacher quality* lacks a clear definition (Kennedy, 2008). Any discussion regarding the quality of teaching must first define this term. Defining teacher quality is perhaps one of the more ambiguous aspects of educational rhetoric, as it pertains to VAM. The term *teacher quality* can be used to describe a variety of teacher-related items, such as teachers’ classroom practices and credentials. For instance, those interested in teacher recruitment may use teacher quality to refer to tested ability. These differences often cause confusion and make it difficult to define or ascribe specific characteristics associated with high-quality teaching. Methods of defining *teacher quality* need to have a sound and defensible conceptual basis, especially if they are to be used in quality assurance decisions such as employment, promotion and professional certification, as many states and districts are already doing (Ingvarson & Rowe, 2008). Differences in definitions combined with differences in measuring teacher quality can produce contradictory findings about educational efficacy, potentially undermining the process of teacher evaluation (Goe & Stickler, 2008).

Considering the high-stakes decisions that are often made as a result of analyzing teacher’s VAS, selecting a valid VAM model to use when calculating teacher’s scores is vital. In the literature, *teacher quality* is often synonymous with teacher performance, teacher effectiveness and teacher efficacy. *Teacher Quality* is the phrase that is most commonly used in educational literature. However, an important distinction needs to be made between *teacher quality* and *teaching quality*. The former implies that the quality of instruction lies within the person, whereas the latter implies that the quality of instruction is dependent upon certain skills or methods that are observable, measurable and can be fostered and taught. This distinction is particularly important because it may be nearly impossible to change *teacher quality*, but there is a good chance that *teaching quality* can be measured and changed. For clarification, the term *teaching quality* will be used throughout the remainder of this manuscript. If student performance differences can be substantiated and causally linked to specific characteristics of teachers, the potential for improvement of education could be great (McCaffrey et al., 2004). There are multiple sources of influence on students’ literacy development and it is assumed that teaching and instructional quality are important sources of that influence. Therefore,
better understanding the aspects of teaching that lead to stronger student outcomes (as evidenced by teachers’ value-added scores) may aid in our understanding of how to train teachers to be more effective and to set clearer expectations for what constitutes high quality teaching.

**Measuring Teaching Quality**

There is accumulating research evidence to suggest that teachers’ credentials, experience, and years of education may make a difference in children’s achievement (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002). However, there are many different teaching qualities to assess and measure with regard to calculating VAS. This study considers teacher knowledge (TK) and teaching experience (total years spent teaching) as factors for inclusion in three VAM models.

Given the social and economic importance of teaching quality at both the national and individual levels, VAM should produce VAS that are accurate representations of teaching effectiveness (Hughes, 2007; Masters, 2004). With many different competing VAM available, this study attempts to determine which model(s) are valid measures of teaching quality by comparing their results with other established measures of teaching quality. This study proposes three research questions: (1). How consistently do 3 VAM predict individual teachers’ TQ (as deemed by VAS), the SFEM (Tekwe et al., 2004), the CAM (AIR, 2011), or the HLMM (Tekwe et al., 2004) when the student outcomes used are closely aligned with one construct, literacy; (2). To what extent do classroom-wide student characteristics & other variables not within the teachers’ control contribute to VAS?, and (3). How well do the results of these models identify teachers who teach reading and vocabulary in ways that are identified as high quality using other metrics, such as teacher knowledge and classroom observation?

In the following chapter, the available literature will be reviewed regarding models used to calculate VAS, along with descriptions of the teacher and student, and school-level variables included in this study.
CHAPTER TWO
REVIEW OF THE LITERATURE

More than one-third of children in the United States lack fundamental reading skills (NAEP, 2005). It is difficult to imagine reading proficiency as an entity unrelated to teaching quality. Quality teaching is an essential part of any educational system. Without high-quality instruction, students may not be provided the opportunity to reach their full potential both in school and in life. The methods used to evaluate teacher quality deserve scrutiny, as this lies at the center of the educational improvement debate. Schools share one common goal, to educate their students, and this goal is accomplished through teachers. Without high-quality teaching, students may not learn as well or succeed academically. The basic literacy skills that children learn in elementary school are the building blocks on which academic, occupational, and social success depend (Spira et al., 2005). The ambitious goal of raising the quality of learning for all will not be achieved unless all learners receive high-quality teaching (OECD, 2001, 2005). Given that teachers are one of the most valuable resources available to schools and students, this study attempts to determine the validity of three VAM and their ability to identify teaching quality.

Value-added methods are quickly becoming education’s answer to the question of how to efficiently and effectively determine teaching quality. However, several key problems exist when considering the implementation of such a system. First, a number of statistical models are available for calculating teachers’ value-added scores. This is problematic when considering which factors are included and excluded in model calculation as well as the assumptions present in each model, providing unique implications for score interpretation and practice. Second, determining the validity of the different models requires an examination of each model with other established measures of teaching quality, allowing for comparisons to be made. Third, it is important to keep in mind that although established, agreed-upon measures for student evaluation exist at the elementary level for literacy, this may not be true for teachers who teach in grades beyond 5th and for subjects that are not reading or mathematics.
Value-Added Methods

Several studies have found the impact of teaching quality as evidenced by student achievement (Sanders & Rivers, 1996; Darling-Hammond, 2000). According to these data, teacher quality is the single most important feature of a school that drives student achievement (Haskins & Loeb, 2007). Therefore, the quality, or lack thereof of a classroom teacher has a direct impact on student learning, and consequently, student achievement. Within the education community, there is a growing trend to use changes in student test-scores to evaluate teaching quality. This approach, known as value-added methods (VAM) is gaining popularity among policy-makers as a way of satisfying accountability demands. VAM are currently being used to assess both school and teacher effects. The focus of this research is on the use of VAM as a means to determine teaching quality (teacher effects), as measured by value-added scores. This study seeks to determine which value-added model(s) are valid measures of TQ.

Stemming from the production function literature in economics (Priester & Wang, 2010), value-added methods were originally used in the business sector to evaluate the financial effectiveness of companies. However, the evaluation of fiscal fitness has been transferred to the realm of education (via VAM) and is currently being used in the analysis of teacher effects. Value-added models were designed to measure change in student achievement, adjusting for differences in students’ prior achievement, home, and social background (Rowan et al., 2002). This approach averages data for every teacher and compares the growth the students of those teachers make with the growth made by other students in the same grade and/or subject. These data may then be adjusted according to the trajectory of every student, providing a platform for evaluating the effect of teachers on student learning (Haycock & Crawford, 2008).

Benefits of Value-Added Methods

In 2012, the era of education’s reliance on Value-Added Methods has arrived. These methods attempt to isolate the contribution teachers have on student learning conditional with certain individual student characteristics such as background and academic preparation. Teachers are characterized as high quality if their students make above average improvements in academic achievement relative to other teachers with comparable students. Advocates of this method state that it is a much better means of
determining teaching quality as compared to traditional test measures. Sanders (1997), an early proponent of VAM, suggested that a statistical method for measuring the influence of teachers on student learning that focuses on student improvement rather than absolute scores (change versus status scores) is, the “only fair, reasonable thing to do if you are going to have an accountability system”. According to advocates of VAM, one of its greatest strengths is that student’s initial achievement levels are taken into consideration when factoring the added-value classroom teachers contribute to student learning. This may help to offset those students who begin the academic year above grade-level, as well as those who begin below grade level, in an attempt to determine how much a student has learned as a result of a particular teacher (Rothman, 2010).

Arguably, the quality of classroom instruction provided by effective teachers is one of the most important aspects of education, because students who do not receive high-quality instruction will not have the opportunity to learn and reach their full academic potential. Research continues to document the fact that although many factors contribute to children’s success in school, their teachers’ capacity to teach effectively is among the most important (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002; Nye, Konstantopoulos, & Hedges, 2004). However, in the past, identifying teaching quality has been fraught with controversy and continues to be an important area of debate and focus in educational research today. This study attempts to discern which model(s) most accurately predict teaching quality.

**Limitations of Value-Added Methods**

The value-added methodology is not without its limitations, chief among them is the challenge of what to do with teachers who teach in grades or subjects not tested by standardized measures namely Social Studies, History, and the Humanities (including art, music, physical education, etc.). However, this problem is rectified if accountability measures are not required for these subjects. Or, (as is currently the case in Florida and many other states), standardized tests are being drafted for any subject that does not have an assessment. Another potential problem relates to the uncertainty associated with the calculation of value-added scores. Like any statistical measure, VAS are subject to measurement error, and it can be challenging to ascribe a precise numerical value to an individual teacher. However, education has long since relied on statistical measures for
many calculations, such as FCAT-score analyses for assigning school grades (i.e., report card systems), among others.

Secondly, measurement error can distort regressions in residual score models, such as those calculated with Hierarchical Linear Modeling (HLM, Raudenbush, Bryk, & Congdon, 2000) and Hierarchical Linear Methods Model (HLMM, Tekwe et al., 2004; McCaffrey, 2004). Measurement error can also lead to instability in teacher and school effects (Linn, 2000). Equating error may lead to instability in estimating teacher and school effects as well (Arce-Ferrer et al., 2002; Haertel, 1986; Waltman, 1997). In addition, measurement error may lead to differences in score reliability for gain score models, including the Simple Fixed Effects Model (SFEM, Tekwe et al., 2004). Often, measurement error is much larger at high and low scores and smaller at middle scores; which might have a substantial effect on both teacher and school effect estimates (Kolen, 2008). Although measurement error can occur during any statistical calculation, and may be a valid concern during certain instances, careful data entry, use of an updated statistical software program, multiple measures of the same construction, and thorough examination of the output can minimize most measurement errors, but they can never fully be eliminated.

Third, it is worth noting that within the elementary education domain (K-5th), there are accepted standardized assessments for measuring student outcomes, particularly for literacy skills, such as the WJII Tests of Achievement, among others. However, this agreement occurs far less often in middle and high school, where certain subjects (namely social studies, science, and the humanities, including art, music, and P.E.) either do not have standardized assessments or there is no consensus regarding which evaluations to implement.

According to recent research conducted by Sass (2008), teachers’ value-added scores may fluctuate widely from year to year, pointing to the fact that they may not be stable indicators of teaching quality. This calls into question the practice of using value-added measures to determine teaching quality. Conversely, these findings could indicate the important roles factors such as classroom composition or student characteristics play in determining a teacher’s value-added score, potentially explaining some of the year-to-year variance in VAS. Therefore, these variables (classroom composition and/or student
characteristics) may be a necessary component when calculating VAS for teachers. Classroom composition, and student characteristics and their implications for teacher’s value-added scores will be discussed in detail in the methods chapter.

In addition to concerns regarding the instability of test-score data, some researchers provide further reasons for caution in the use of student test-scores as a single metric of teaching effectiveness. For instance, unreliable standardized tests can undermine the accuracy of value-added measures or those that cover only some state and district learning objectives (i.e., those that are not well-aligned with the curriculum). Estimated teacher effects have been shown to depend on skills that are measured by achievement tests (Linn, 2000). Therefore, test content likely has a substantial influence on school and teacher effects estimated from both gain and residual score models (Kolen, 2008). These limitations may cause doubt regarding the use of VAM to assess individual teacher effectiveness, or even its ability to provide reliable estimates of student learning growth (Berry, 2010). Furthermore, it is important to keep in mind (even if VAM prove effective for determining teaching quality) that student achievement tests were not designed to assess the quality of teaching, but rather to assess the content area knowledge students’ possess. Haskins and Loeb (2007) caution that value-added models can help identify effective teachers, but that this should only be part of the equation. These researchers suggest that test-score changes should not be the only factor in a system of evaluating teacher performance. They suggest instead that policymakers not be seduced by the prospects of relying solely on standardized test results as a means to determine who teaches effectively, and consider VAM as one of several evaluation tools to determine teacher quality. These researchers propose instead that school systems evaluate teachers with a comprehensive performance review system based on a combination of student gains, principal evaluations, parent evaluations and perhaps other measures, using a procedure developed cooperatively by schools administrators, teachers, and others, including parents (Haskins & Loeb, 2007). However, including additional measures for evaluating quality teaching can be timely and costly, hence the allure and potential benefit of using VAM. While today’s VAM for evaluating teaching quality can provide useful information, some researchers caution that they may not be reliable measures for making high stakes decisions. However, the rationale behind value-added
methods is to hold teachers accountable for the learning gains of the students they serve (Raudenbush, 2004), assuming the resulting scores are fair representations of teaching quality.

In defining teacher effectiveness, new evaluation tools (or alternative VAM calculations) may be needed to measure student achievement to determine teaching quality. To some, teaching may be far too complex to be judged by a single metric. Accumulating research evidence calls for a range of tools for measuring teacher effectiveness based on evidence of the following: (1.) Student learning, including evidence drawn from classroom assessments and value-added student achievement scores; (2.) Teacher performance (via measures other than VAM); and (3.) Measures of teacher knowledge, skills and practices associated with student learning (Berry, 2010; Haskins & Loeb, 2007). Most VAM formulas take these factors into consideration when calculating a VAS. Value-added methods of teacher evaluation may be able to improve our ability to identify effective teachers with regard to their effect on student achievement, but some suggest they may need to be used in conjunction with other methods. More research is needed to determine which model(s) of measuring teaching quality is most in-line with established measures of measuring TQ. This study investigates multiple VAM models using students’ literacy scores to determine their effectiveness in identifying teaching quality.

Modeling Value-Added

Three of the models used to calculate value-added scores for teachers are the Hierarchical Linear Method Model (HLMM, Tekwe et al., 2004), the Covariate Adjustment Model (CAM, AIR, 2011) and the Simple Fixed Effects Model (SFEM, Tewke et al., 2004). A popular approach to value-added measures has relied on HLM analysis (Goldstein, 1997, Phillips & Adcock, 1996), to which these three models belong. In these hierarchical regression models, the focus is on differences between student test scores in current year and previous years (hierarchical growth models that focus on estimating growth trajectories). The CAM, (Florida’s official adopted VAS model) takes into consideration the academic achievement level of students in the fall (as a starting point), whereas the HLMM model does not. The HLMM models that have been studied in the literature are special cases of the general mixed effects model. In the HLMM,
schools are assumed to be a random sample from a larger population of schools. The SFEM model differs from the HLMM model, where school populations are considered fixed. In gain score models (SFEM), the focus is on differences between predicted and observed scores.

In addition to the three methods mentioned above, the Layered Mixed Effects Method (LMEM) is another method available, however not included for the purposes of this study, which was designed to estimate school effects on student learning gains (Sanders and Horn, 1994). This model provides the foundation of the Tennessee Value-Added Assessment System (TVAAS). The TVAAS was the first large-scale implementation of VAM for determining school effectiveness. The LMEM includes neither a direct measure of gain nor a measure of incoming knowledge as a covariate. It, does, however produce value-added measures of school effects by utilizing the information in non-zero covariance between test scores at different times (Sanders and Horn, 1994). Several researchers have demonstrated that the LMEM can be viewed as a model for change scores with random school effects (Carter et al., 2001 & McCaffrey et al., 2003). In addition, a LMEM can be specified to either analyze multiple subject area test scores simultaneously (multivariate LMEM) or separately (i.e., univariate LMEM). However, Tekwe et al. (2004) found that when compared to the SFEM and the HLMM, the LMEM results were highly correlated to both; yielding almost identical results as the SFEM. Therefore, when choosing between the LMEM and the SFEM, the SFEM is preferred for parsimony, hence its inclusion in the current study’s analyses.

There is a natural desire on the part of the public and education professionals for school accountability systems to involve methods easily understood by many, not just those with extensive methodological training (Tekwe et al., 2004). The Simple Fixed Effects Method model does just that by analyzing student test score changes (gain score analysis). The virtue of this model is its simplicity. However, a potential shortcoming of this method is that the SFEM ignores possibly confounding student and school-level factors, such as minority status (race) and poverty (Free and Reduced Lunch, (FRL) status), which might unfairly bias comparisons among teachers.

Tekwe et al. (2004) compared four models of VAM: 1. HLMM (simple unadjusted change score), 2. HLMM (adjusted for demographic and intake score with
outcome defined by change score), 3. LMEM (a multivariate, layered, mixed-effects model) and 4. SFEM (a simple, fixed effects model, gain score approach). In the study, the authors found no significant differences between the SFEM and the LMEM models. This finding indicates SFEM (the simplest model) is preferred over the much more complex LMEM. These models produced results that were highly correlated ($r > 0.91$). Thus, there is little benefit in using the more complex model (LMEM). However, the authors found significant variability between the SFEM and the HLMM (adjusted) models (correlations ranging from 0.60 to 0.90). These findings warrant closer investigation. The main difference between the SFEM and the HLMM models is the exclusion or inclusion of student level and school level variables. In the SFEM, student and school level variables are not included in the analyses. In the HLMM model, student and school level variables are included. The SFEM has been criticized for not including these student and school level variables. Critics of the SFEM also argue that formal education does not occur in a vacuum, and there are many factors potentially influencing student achievement aside from instruction, including personal propensities, resources, physical and mental maturation, home environment, cultural heritage, institutional and community resources, among others (Kupermintz, 2003). Alternatively, several studies have found similar results (comparable VAM scores) when calculating models including and then recalculating models excluding these student and school-level variables, indicating that perhaps these factors are not quite as influential as may initially appear. These mixed results indicate a need to determine if multiple models of VAM produce valid measures of teaching quality (by identifying the same teachers), hence the inclusion of both the SFEM and HLM models in this study.

In the spring of 2011, the state of Florida enlisted the help of a diverse group of stakeholders (teachers, administrators, and others), known as the Student Growth Implementation Committee (SGIC) to make final recommendations for a specific value-added model for statewide use. The SGIC chose Florida’s official adopted model, a Covariate Adjustment Model (CAM), which uses prior student test scores as variables predicting teachers’ VAS. First, student-level prior test scores and the covariates are used to establish statewide conditional expectation. This expectation is the score a student is expected to have, given his/her prior test score history and measured
characteristics. Unique to this model is the inclusion of a common school component, an indicator of the amount of learning that is typical for students in each school that differs from the statewide conditional expectation. By employing three value-added methods (SFEM, HLM, and CAM), this study seeks to determine the extent to which these different models’ VAS are correlated with one another, to establish whether their resulting value-added scores identify the same teachers.

It is widely accepted among educators and researchers that value-added assessment of teachers is preferred over assessments based on status-scores alone. However, a theory of action does not exist for VAM, that is, it is not yet evident in the research that VAM aid in improving the quality of teaching. What remains unclear is which model is preferred when calculating these value-added scores. According to Ballou et al., (2004), (an advocate of SFEM), school and student level variables do not seem to have an impact on value-added scores and therefore should not be included in a statistical model used to calculate such scores. However, without fitting several models with the same data, an impartial comparison cannot be made to determine if the measures are valid indicators of teaching quality. This study intends to provide useful information to those who must choose between competing methods for calculating VAS, for determining TQ by using aggregated teacher data in regression analyses. Alternative methods employed by states with accountability systems include the use of regression analyses with aggregated data (Richter and Brorsen, 2006). More than 35 states currently have comprehensive report card systems detailing a variety of issues including test scores and a comparison of school variables with district and state averages. Typically, regressions are calculated for the purposes of obtaining VAS (Sanders and Horn, 1994). Regression analysis is a method that may help determine student improvement and teacher effectiveness, and is also a model that is more readily understood. Multilevel analysis has been criticized for being a complicated statistical analysis that school officials may have difficulty understanding, thus the need for a more stream-lined approach for answering the question of teaching quality, such as linear regression based on aggregated data (Ladd 1996).

This study relies on the use of aggregated data for answering the second research question with regression analysis. Although commonly used, the use of aggregated data
is not without its limitations (Richter and Brorsen, 2006). The most common complaint voiced by those opposed to the use of aggregate data is that the effects upon individual students are unavailable. In addition, aggregate data does not allow researchers to study differential effectiveness, which distinguishes between teachers that are effective for low achieving students and teachers that are effective for high achieving students. However, Richter and Brorsen (2006) show that the criticisms of aggregate models have been overstated and advocate for the use of aggregate data when disaggregate data are unavailable or when maximum likelihood estimators are employed.

**What Contributes to Teacher’s Value-Added Scores?**

**Teacher-Level Factors**

Teacher’s value-added scores are reflections (at least in part) of the knowledge they possess, the experiences they bring to the classroom, the training they receive, the classroom environment they foster, the students in their classroom, and the learning environment the teacher and students create jointly, among others. A variety of teacher-level variables are used for calculating value-added scores for determining teaching quality, including, content-area knowledge, race (ethnicity), years of experience, gender, educational attainment, and professional development attendance. For the purposes of this study, the following teacher-level factors will be examined in detail in relation to assessing the quality of instruction in classrooms: teacher gender, teacher ethnicity, years of experience, content-area knowledge, and overall teaching quality (via fidelity rubric data). Descriptions of the teacher-level variables included in this study are provided below.

**Teaching Experience**

The relationship between teaching experience and student achievement receives considerable attention in the empirical literature, with somewhat mixed results. Teaching experience (as defined by total number of years spent teaching) often has monetary consequences for policy makers and administrators as teacher tenure has long been the basis of teacher compensation rather than student achievement (Munoz & Chang, 2008). The logic behind this practice is that the more years of experience a teacher has, the more effective the teacher should be in teaching. Several studies support this claim, including, Ost, (2009) and Nye et al. (2004), who found statistically significant positive results (via
increased student achievement) for teacher experience in second grade reading. In addition, Huang & Moon (2009) found that teachers who had more than 5 years of total teaching experience had a positive association with increased student achievement. Huang and Moon (2009) concluded that the effect of seasoned teachers (those who had more than 5 years of experience teaching at a particular grade level) is meaningful in comparison with the other effect sizes found and apt to affect student achievement positively. Given that teacher effects are cumulative, and effects are similar across grades, students taught by seasoned grade-level teachers for 4 years in a row may score more than one standard deviation higher, or approximately one grade level more than students taught by beginning (inexperienced) teachers. This difference can have a profound impact on the academic success of students.

Teaching experience may exhibit nonlinear effects, with teachers becoming more effective in the first few years of teaching with learning gains decreasing or tapering off (Goldhaber, 2008; Kane et al., 2006). Huang and Moon (2009) found teachers constantly improved teaching effectiveness until the 21st year and declined beyond that. In their study, the most effective teachers had 19-24 years of experience at grade level, and were associated with increased student reading achievement. The authors conclude that their results showed that teachers improved their practice over time, but learning growth slowed over the years, with the largest growth coming from the first few years of teaching (2009). These studies echo the plateau effect for teaching experience beyond the first 5 years that is well documented and evident in the literature.

Even though many researchers have found positive effects for teaching experience as it relates to student achievement, recent research suggests quite the opposite. Connor et al. (2008) considered teachers who teach ineffectively year after year. According to this research, more years of experience will never lead to improved student outcomes. In addition, several other studies do not detect meaningful differences between more and less experiences teachers (Carr, 2006; Gallagher, 2004; Harbison & Hanushek, 1992). Goldhaber (2002) concluded that teaching experience does not consistently and positively influence student learning. This conflicting evidence calls into question the use of teaching experience as a predictive variable in VAM calculations, and is why the present
study includes teaching experience as a variable, to determine if this is a necessary factor for calculating VAS.

There is a substantial body of work on other methods of evaluating the quality of teaching, here referred to as "established measures". These include teachers' specialized knowledge (content-area knowledge), specific classroom practices (overall teaching quality, TQ), and teacher’s perception of their student’s social skills, among others, which are discussed below.

**Teacher Knowledge**

Teacher’s content-area knowledge is broadly defined as the professional skills that a teacher brings to the classroom. The available literature on teacher knowledge varies greatly in how this term is defined, and how this knowledge is determined. Teacher knowledge can refer to both pedagogical knowledge and content area knowledge. Pedagogical knowledge is also known as pedagogical competence and simply refers to what a teacher can do in a classroom, i.e., what he or she demonstrates through instructional delivery. Determining pedagogical knowledge can be a particularly subjective endeavor, and was thus not included in this study. Content area knowledge, also known as subject matter knowledge, refers to what a teacher knows about his or her particular topic. A common approach to determine teacher knowledge is through the use of surveys or multiple-choice tests. Content-area knowledge is included in this study as a variable for teacher knowledge.

In line with current legislation (NCLB), effective (high-quality) teachers must understand the concepts to be taught while possessing the specialized knowledge to teach these concepts successfully (Piasta et al., 2009). Thus, content-area knowledge may play an important role impacting teachers’ ability to effectively instruct their students.

**Quality of Teaching (Fidelity Rubrics)**

In line with several other recent studies’ use of observational measures (Pianta et al., 2008; Carlisle et al., in press) for determining teaching efficacy, this study uses a fidelity rubric to assess the quality of teaching. In this study, fidelity rubrics were developed and implemented as an observational checklist for classroom practices such as organization, responsiveness, and warmth, among others. Similarly, Pianta and colleagues developed the Classroom Assessment Scoring System (CLASS), a
theoretically based and empirically supported observation instrument designed to assess the quality of interactions between teachers and students in the classroom. The CLASS measures three broad domains of teacher-student interactions, including emotional support, classroom organization, and instructional support. For implementation purposes, trained assessors observe classrooms and use a detailed manual to assign scores based on a 7-point scale (low- 1-2, mid-range-3,4,5, & high-7).

The CLASS framework is organized according to three categories of teacher-student interaction, emotional support, classroom organization, and instructional support. This three-latent factor structure has been validated in just under 4,000 preschool to fifth grade classrooms (Hamre et al., 2007), suggesting that this organization is a useful and valid conceptualization of teacher-student interactions that occurs in elementary classrooms. Each of the three latent domains include several key dimensions that have empirical support for effects related to child outcomes. The fidelity rubric employed in the current study includes an organization and planning category, and a warmth/responsiveness category, similar to the classroom organization and emotional support categories included in the CLASS.

In a recent study comparing first-grade teachers’ responses to professional development programs in reading, Carlisle et al. (in press), developed a post-observation survey used to evaluate the extent of impact the PD had on teachers’ instructional practices. In the study, trained researchers conducted systematic classroom observations of the dedicated literacy block for each teacher. The following key features of the instruction were included in the post-observation survey: purpose of instruction, modality, grouping arrangements, adults present in the classroom, and material(s) used during instruction. The validity of the fields coded in the observation protocol was determined by reading research on the components and features of effective literacy instruction in first grade (Snow et al., 2005, US Department of Education, 2002). The fidelity rubric used in the current study mirrors this observation survey because both include a focus on grouping for instruction.

**Teacher Perception of Student Social Skills**

The social and academic skills that students bring to the classroom likely have an impact on their academic achievement. With regard to social skills, the Social Skills
Rating System (SSRS, Gresham & Elliott, 1990) is used in the current study as a composite measure of teacher’s perceptions of their students’ characteristics. The SSRS measures students’ abilities in three distinct constructs, including Academic Competence (AC), Problem Behaviors (PB), and Social Skills (SS). AC is defined as a teacher’s impression of a student’s overall academic functioning, including his or her skills in comparison with other students, his or her motivation to succeed, his or her classroom behavior, and parental support. PB are any behaviors that interfere with the acquisition or performance of important social skills (Gresham & Elliott, 1990) and include potentially negative behavior, such as externalizing and internalizing behavior problems and hyperactivity. SS are defined as socially acceptable learned behaviors that enable a person to interact effectively with others and to avoid socially unacceptable responses (Gresham & Elliott, 1990). Research suggests that there is a strong relationship between social behavior and academic achievement (DiPerna & Elliott, 1999), hence the use of the SSRS in the current study as a way to gauge student characteristics. A recent study highlights the use of the SSRS for understanding student achievement in relation to student characteristics.

Malecki & Elliott, (2002) investigated the relationships between students’ social skills, problem behaviors, academic competence and academic achievement. Their results indicated (a) social skills are positively predictive of concurrent levels of academic achievement and (b) problem behaviors are negatively predictive of concurrent academic achievement. In addition, the authors note that social skills mediate students’ ability to function academically. Overall, the research seems to support elementary school teachers as valid and accurate judges of students’ achievement, albeit with some biases (Tani & Connor, 2011).

Several studies support the accuracy of teacher’s perceptions of student’s academic abilities. For example, Demaray & Elliott (1998) investigated the accuracy of teachers’ judgments of students’ academic achievement and found that teacher’s judgments of students’ academic achievement on the Academic Competence scale were correlated moderately high ($r=.70$) with students’ actual achievement scores. This research suggests that teachers’ judgments of student performance may be as accurate as standardized achievement tests. The results from this study indicate that teachers can
provide highly accurate characterizations of students’ academic achievement. The authors’ conclusion supports the validity of teachers’ judgment of academic achievement. The aforementioned measures of evaluating teaching quality are included in this study's methods section, which are discussed in the following chapter.

**School-Level Factors**

Several school-level factors are available for use in VAM formulas, including SES, parent-education level, and others. For the purposes of this study and the calculation of VAS, the percentage of students in each school who were eligible to receive free and reduced priced lunch (FRL) was used as an indicator of school-level socio-economic status (SES).

**Student-Level Factors**

Student-level factors are also likely to have an impact on teacher’s value-added scores. For the purposes of this study, the following two student-level factors will be included in the VAM models: characteristics of the students in the classroom (classroom composition, i.e., classroom mean self-regulation), and student characteristics (social skills, the academic skills at the beginning of the school year).

*Classroom Composition*

The composition of children in a given classroom, particularly with regard to their ability to self-regulate behavior and attention, is apt to have an effect on the teaching and learning environment, affecting a teacher’s VAS. The concept of self-regulation is helpful in aiding our understanding of group dynamics, in this case, classroom-learning environments. Evidence suggests that, in addition to early literacy skills, individual differences in students’ self-regulatory skills contribute to differences in academic achievement (Smith, Borkowski, & Whitman, 2008). Self-regulation refers to the strategies and processes that allow children to pursue goals by maintaining or altering internal states and responses (McCabe et al., 2004), Posner & Rothbart, 2000), and, at least for younger children, allows them to plan and manage their behavior (Ponitz et al., 2008). Self-regulation underlies multiple skill domains related to controlling and directing behavior, and enables students to function in cognitively challenging settings, such as the second grade classrooms in this study. Both literacy and self-regulatory competence appear to be associated with teacher practices, such as planning, organizing,
and managing the classroom environment (Cameron et al., 2008; McClelland et al., 2007; Ponitz et al., 2008). Behavioral self-regulation has been linked to academic achievement in kindergartners (Howse et al., 2003) and early elementary aged children (Howse et al., 2003).

Student self-regulation has unique implications for achievement gains, as the classroom composition of low and high self-regulation will inevitably have an effect on the classroom-learning environment, and consequently, student performance (Skibbe et al, in press). Self-regulation is an important skill to consider when assessing children’s early literacy development in particular. Children who exhibit better work-related skills, including self-regulation, have significantly better reading-related skills in kindergarten and second grade (McClelland et al., 2000). Considering that effective teachers may promote literacy skills by encouraging their students to develop self-regulation (Paris & Paris, 2001; Pressley et al., 2001), this is an important component for calculating value-added scores.

Self-regulation is generally defined as the deliberate modulation of one's responses to stimuli, which includes how an individual functions in the face of different types of activation (e.g., attentional, behavioral, or emotional activation; Baumeister & Vohs, 2004; Calkins, 2007). Under this definition, self-regulation includes a specific constellation of skills critical for persisting on academic tasks and completing work independently (Blair & Razza, 2007; Deater-Deckard, Petrill, Thompson, & DeThorne, 2006; McClelland, Acock, & Morrison, 2006; McClelland et al., 2007). Students with stronger levels of overall self-regulation, measured with tasks that necessitate integrating multiple component skills, (specifically attention, working memory, and inhibitory control), generally achieve at higher levels compared to students with weaker overall self-regulation (Howse, Calkins, Anastopoulos, Keane, & Shelton, 2003; McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009).

In this study, greater self-regulation supports increased academic success (specifically literacy skills) and builds on the conceptualization of self-regulation as the capacity to moderate attention and behavior in response to contextual contingencies (Connor et al., 2010). In particular, students with stronger self-regulatory ability might be better able to focus cognitive resources (e.g., sustained and flexible attention, working
memory) on the learning task at hand (e.g., listening to teacher instructions, reading text, completing a writing activity). Compared to poorly regulated peers, a student with stronger self-regulation might be more likely to ignore distractions posed by other students and unrelated classroom stimuli. In the context of a classroom where students are expected to manage their own attention (e.g., in independent learning centers, in transitions between and during small and large group instructional times), students with better self-regulation may be more likely to maximize their learning opportunities throughout the day. In theory, this capability to mobilize and maintain cognitive attention and engagement would lead these students to experience more, or at least more efficient, time-on-task with the literacy activities made available to them within the classroom. This efficiency could ultimately lead to more efficient learning of literacy competencies, such as reading comprehension and vocabulary, two key measures of literacy skills. Key to this model is the assumption that specific teacher practices (planning, classroom management, use of routines, and providing opportunities for students to work independently and with peers) would tend to support the development of stronger self-regulation and, hence, stronger academic achievement, such as literacy gains. The composition of students in a classroom (as measured by their self-regulation skills) may not only alter the learning environment, but may also influence the teaching quality. Thus, classroom composition and its effect on teaching and learning can be seen as a symbiotic relationship to instructional practice.

**Difference from Modal Age**

This student-level variable is used as a marker of retention. This variable is determined by calculating the mean age of students per classroom, and subtracting each student’s age from that average age, thus providing an indication of students who are older, or younger than expected per grade.

As discussed, direct, valid and reliable assessment of students’ academic skills is crucial when computing value added scores inasmuch as students’ fall reading scores, for example, are the single best predictor of their spring scores (Connor et al., 2004). In this study, the Woodcock Johnson III test of achievement (Mather & Woodcock, 2001) (a standardized battery) was used to assess student’s reading skills. Specifically, the passage comprehension (PC) and the picture vocabulary (PV) subtests were used to assess
student’s reading comprehension and vocabulary skills. Passage comprehension was selected as a child-level outcome in this study as oral reading fluency measures are typically considered indicators of reading competency at the elementary level (Fuchs et al., 2001). Picture vocabulary was chosen for use in this study as research indicates that early vocabulary knowledge predicts future reading success (Senechal et al., 2006). Reading fluency often precedes successful reading comprehension, and develops alongside oral language skills, such as expressive vocabulary.

A combination of factors are likely to contribute to a teacher’s classroom environment, including the above-mentioned teacher, student, school, and classroom-level variables. These variables create a unique learning environment for students and a distinctive instructional environment for teachers, which may impact their VAS.
CHAPTER THREE
METHODOLOGY

Purpose

The purpose of the proposed study is threefold: (1.) to compare the rankings of teachers using three different value-added models; (2.) to examine factors that are likely to contribute to teacher’s value-added scores; and (3.) to determine how well the value-added scores are related to other non-value-added established measures generally accepted to be representative of higher quality literacy teaching. Much research has been conducted regarding value-added methods, including comparisons of various models. However to date, there is much less attention to sources of influence on value-added scores, and, to the best of this researchers’ knowledge, no one has attempted to ascertain what the value-added scores explain in relation to other non-value-added markers measuring teaching quality. Variance in teaching quality exists, but if the factors that contribute to teacher’s value-added scores are isolated, it would then be possible to recruit, and train or foster individuals with these characteristics to enter and remain in the teaching profession.

This study compared second grade teacher’s value-added scores by fitting three statistical models with the same data and comparing their resulting value-added scores. Then teacher’s value-added scores were examined to determine the extent of agreement among models and their VAS. Finally, the relationships between value-added scores (produced from three different models) and several other measures of teaching quality were examined to determine how well value added scores predict other well-regarded indicators of teaching quality, including teacher practice and teacher knowledge.

Research Questions

The present study was designed to address three research questions:

1. How consistently do three VAM predict individual teachers’ teaching quality (as deemed by VAS), the Simple Fixed Effects Model (SFEM, Tekwe et al., 2004) the Covariate Adjustment Model (AIR, 2011), and the Hierarchical Linear Methods Model (HLMM, Tekwe, et al., 2004) when using student outcomes that are strongly tied to one construct, literacy.
2. To what extent do classroom-wide student characteristics (i.e., classroom composition), and other teacher variables contribute to teachers’ value-added scores?

3. How well do the results of these models identify teachers who teach in ways that are identified as high quality using other metrics (external measures of TQ)? That is, how well do value-added scores agree with other measures of teaching quality. These external measures of teaching quality include teacher knowledge, classroom literacy instruction, and general quality of the classroom-learning environment, assessed using teacher fidelity rubrics.

**Study Design**

This study utilized a longitudinal correlational design. Part of a larger randomized control study, teachers in this study were randomly assigned (within schools) to receive training in one of two interventions: (1) the Individualizing Student Instruction in Reading intervention (Connor et al., 2007); or (2) the Robust Vocabulary Intervention (Beck et al., 2002). This study takes advantage of this design because all teachers received training and the extent to which they participated in the training was measured, all teachers were observed in the classroom during literacy instruction, and their students’ literacy skills were assessed before and after implementation of the interventions (pre-post design). The results of the larger second grade study are presented in a published chapter and in a manuscript currently under review (Connor, 2011; Connor et al., 2011).

**Participants & Setting**

The participants in this study included 40 second-grade teachers and their students \(n=647\) in 8 northeast Florida elementary schools during the 2007-2008 academic year. The student sample contains 49% male and 51% female students. In addition, the sample is ethnically diverse, with 49% Black, 38% White, 5% Multiracial, 4% Hispanic, and 4% Asian. These second grade students and their teachers were selected because a majority of the current research focuses on either first or third grade. Second grade was chosen as it is an important academic year for students. During second grade, students are expected to acquire reading skills beyond simple decoding and demonstrate competency in literacy by reading fluently and with understanding. Typically, students acquire the skills necessary to read during first grade (learning to read), and by third grade, are well on their way to becoming fluent comprehenders (reading to learn). Moreover, students who
fail to achieve grade-level literacy skills by the end of second grade are unlikely to do so during their school career (Spira et al., 2005). Therefore, it appears as though second grade is a year when many essential skills are learned, enabling students to read proficiently. Hence, second grade was selected as the focus for this study.

**Data Collection Procedures**

**Teacher-Level Measures**

For this study, three types of teacher variables are included: (1) those that describe teachers and will be included in models for computing value-added scores; (2) those that describe aspects of the classroom environment that are likely out of the teachers’ control, such as the number of students with poor self-regulation, but are also likely to impact teachers’ VAS, and (3) variables that are measures of teaching quality that are not VAS. For the latter, the teacher had to be able to manipulate the indicator during the year of the Randomized Control Trial (RCT). Thus, for example, teachers’ years of experience, a widely used metric of teaching quality, is not considered an indicator of teaching quality because the teacher could not control how long he or she had been teaching during the year of the study. Moreover, there is good evidence to suggest that years of experience may not be correlated as expected with student outcomes (Connor, Son, et al., 2005). However, teacher’s specialized knowledge about teaching reading, and specific classroom practices are within the teachers control and should lead to more effective or higher quality teaching (Carlisle et al., in press). These latter variables will be used as indicators of teaching quality.

*Years of Experience*

Data for teachers’ years of experience was collected via the Teacher Background Questionnaire (TBQ). This questionnaire asked teachers to provide their total number of years spent teaching (see Appendix D for TBQ).

*Teacher Knowledge*

In an effort to ascertain teacher’s knowledge of literacy concepts, a researcher-developed survey was administered in the fall. This Teacher Knowledge Survey (TKS) consists of 34 multiple-choice questions and 6 short answer questions (see Appendix C). The questions in the multiple-choice section are a combination of proposed classroom teaching scenarios and straightforward questions about reading and language arts topics.
including phonemes, morphemes, graphemes, phonics, phonemic awareness, and rhyming. The following is an example of a multiple-choice question: “You say to your students, “The word is taught. What word would you have if you said taught without the /t/ sound?” What type of task are you asking for? a). rhyming b). blending c). elision d). none of the above e). I don’t know. The answer is c). elision. An example of a short-answer question is, Name a syllable type and provide an example. All TKS were individually administered and hand scored to ensure accuracy during data entry.

*Teachers’ Perception of Students’ Skills (Academic Competence)*

Teachers’ ability to correctly judge their students’ academic competence may reflect their quality of teaching. Teachers’ perceptions of students’ academic competence were measured with the Social Skills Rating System (SSRS), (Gresham & Elliott, 1990). In this study, an average score was calculated for each teacher using aggregated data (the mean score for Academic Competence score per classroom). The SSRS, published by Pearson is designed to evaluate a broad range of socially validated behaviors that affect teacher-student relationships, peer acceptance, and academic performance, among others. Items on each scale of the SSRS are rated according to perceived frequency and importance. The SSRS contains two rating scales, a social skills scale and a problem behaviors scale. This study includes Academic Competence as a measure of teachers’ ability to accurately judge their students’ academic abilities. Academic competence (AC) is defined as a teacher’s impression of a student’s overall academic functioning, including his or her skills in comparison with other students, his or her motivation to succeed, his or her classroom behavior, and parental support. Reliability for the SSRS has been demonstrated via internal consistency (Academic Competence α reliability = .95). The validity of the SSRS has been found to be strong, with criterion-related, content and construct validity all being documented.

*Quality of Teaching (Fidelity Rubrics)*

In an effort to evaluate the teaching quality provided to the students in this study, teacher fidelity rubrics were created (see Appendix D for rubric). The teacher fidelity rubric was designed to assess the quality of teaching in four key domains, instruction, organization, vocabulary, and warmth/responsiveness. Trained research partners completed this rubric for each teacher, following a protocol designed to establish inter-
rater reliability of at least 85% (kappa > .70). The categories included in the rubric are (1.) individualized instruction, (2.) organization & planning, (3.) vocabulary, and (4.) warmth & responsiveness & control/discipline. Each category was rated on a likert scale ranging from one to six (1-low quality, 6, high quality). A rating of one was given to teachers demonstrating consistently weak implementation. A rating of a three was given to teachers demonstrating characteristics, but inconsistently. A rating of a six was reserved for those teachers who were exemplary in the given area. Fidelity rubrics were completed for each teacher twice per year, once in winter (when instruction is more stable), and once in spring to determine if classroom practice had in fact been altered as a result of study condition. Of note, the quality of teaching (TQ) may change throughout the course of the academic year, as a result of the training the teachers received in this study.

**Classroom-Level Measures**

*Student Self-Regulation*

In the present study, a student measure of self-regulation assessed in the fall was aggregated at the classroom mean to provide an index for the extent to which students entering the teachers’ classroom generally had stronger or weaker self-regulation, with the assumption that the assignment of the students in the classroom is at least partly out of the teachers’ control. In general, the higher proportion of poorly regulated students, the more difficult the class as a whole is to teach (Skibbe et al., in press). Students’ self-regulation was assessed using the complex version of the Head-to-Toes Task (HTT, Cameron, 2008). This self-regulation measure assesses working memory, attention, and inhibitory control. The HTT task requires children to remember and use two rules to respond to behavioral commands. A complex version was used in this study as the original HTT reached ceiling levels in children older than 5 years (Cameron et al., 2008). The complex, expanded version, was used, the Head-Toes-Knees-Shoulders Task (HTKS). The HTKS uses four behavioral commands: “touch your head”, “touch your toes”, “touch your shoulders”, and “touch your knees”. Children are instructed to respond with the opposite response (e.g., touch their toes when told to touch their head). The task is intended to assess self-regulation by requiring the use of three skills: (1) attending to tester’s instructions and commands, (2) using working memory to remember
instructions while processing new commands, and (3) inhibiting natural response to commands (Day et al., in press). The present study seeks to understand how varying degrees of self-regulation affects students’ ability to learn and how this cluster of skills influences the classroom learning environment and teacher’s value-added scores.

**Student-Level Measures**

*Reading and Vocabulary Skills*

One advantage of using the construct of literacy as the student outcome of interest is that there are generally agreed-upon assessments of the construct and of effective ways to teach early literacy (NICHD, 2000, National Reading Panel Report), which is not necessarily the case in other content areas (e.g., science, Southerland, in press). The student data for reading comprehension and vocabulary skills in this study relies on subtests of the Woodcock Johnson III Tests of Achievement (WJIII), (Mather & Woodcock, 2001), which are widely used and highly correlated with many state-mandated achievement tests, such as FCAT (Buck & Torgesen, 2006). Passage comprehension and picture vocabulary subtests were used to assess children’s literacy skills in this study. The subtests of the WJ-III have reliabilities ranging from .81 to .94. The median reliability score on the passage comprehension subtest is .88. The WJIII passage comprehension subtest is an individually administered test of academic achievement. The passage comprehension (PC) subtest assesses the students’ language comprehension and reading skills. The task requires that students orally supply the missing word in a given sentence or paragraph. The items progress in difficulty until the child can no longer supply the missing word(s), or if he or she reaches the ceiling, whichever occurs first. PC is a complex, conceptually driven processing task that measures the ability of reading comprehension as it occurs. As the subject reads, the meaning of the passage is derived through construction of propositional representations based on concepts from stored knowledge. This aspect of the process requires the ability of verbal (printed) language comprehension. Meaning is constructed immediately as the passage is read. As more elements are added to the passage, they are also added to the structure held in immediate awareness via the process of mapping, a central feature of cognition (Ashcraft, 2002; Zhou & Black, 2000). The PC task is then solved through inference (Cain et al., 2004), the process by which the reader determines the referents of
words and ideas, draws connections between concepts (bridging) (Clark, 1977), and derives a conclusion from the passage. In this study, the PC enables an estimation of students’ ability to read. This measure seeks to identify how well students understand what they’re reading and are able to acquire meaning for understanding from text.

Students’ vocabulary skills were assessed via the Woodcock Johnson III picture vocabulary (PV) subtest (Mather & Woodcock, 2001). The median reliability score for the picture vocabulary subtest is .81. The PV test requires the cognitive processes of object recognition, lexical access, and lexical retrieval. This test is a non-reading, lexical-level language development task. Object recognition depends on an analysis of the shape and form of a visual stimulus, although non-shape cues such as color contribute to recognition (Marr, 1982). Lexical access results when representations are activated and spread to semantic attributes of words. Retrieval results when the name of the object is located in the store of lexical knowledge (Gazzaniga et al., 1998). In this task, students are asked to identify the names of objects presented in color on the page displayed in front of them. The items progress in difficulty until the child can no longer identify the object(s) correctly or reaches the ceiling, whichever occurs first.

The SSRS, in conjunction with the HSKT allows a holistic view of students’ social skills, self-regulation ability, and provides a general idea of how these behaviors shape the classroom-learning environment. Additionally, students’ distance from modal age was used as a variable included in the analyses, as a marker of retention.

Data Analysis Procedures

Multi-Level Models

Due to the nested structure of the data used in this study (students nested within classrooms, classrooms nested within schools), this hierarchical structure was accommodated by using multi-level models. Descriptions of each research question include specifications of appropriate analytical models, which involve aggregate data for making inferences about individual teachers in research questions 2 and 3. The strategies that take into account the multilevel features of educational data reflect reality (Burstein, 1980). These strategies also enable researchers to tease out effects from a variety of sources so that they may learn something about the interface of the individuals and the
“groups” to which they belong and the implications of the interface for educational effects (1980).

When analyzing clustered (i.e., nested) data, multi-level models provide numerous advantages over other, more traditional methods of analysis (Ames, in press). A multi-level model enables researchers to obtain statistically efficient estimates of regression coefficients (Goldstein, 2003). By using nested data, multi-level modeling provides more accurate standard errors, confidence intervals, and significance tests, which are generally more conservative than the traditional ones that are obtained by ignoring the presence of clustering (Ames, in press). Additionally, the influence of covariates of higher-level variables on lower level variables can be separated in a multi-level model, allowing for a more detailed analysis to be conducted. Lastly, the relative performance of subjects on the primary outcome variable can be compared in light of the embedded contextual effects of higher-level variables (Yurecko, 2009). For these reasons, HLM was employed to answer the primary research questions in this study. A diagram following the notation in Curan and Bauer (2007) depicting the multi-level model is found, below, in Figure 1.1. This diagram presents a random-intercepts HLMM model, as discussed in Tekwe (2004). Student-level characteristics include gender, ethnicity, HSKT, and years from classroom mean age. Teacher-level characteristics include gender, ethnicity, knowledge, and experience. As can be seen in Figure 1.1, teacher-level characteristics predict the mean per classroom, $\beta_{0j}$, with a random component, $u_{0j}$, representing VAS. These random components are unaccounted for variance in classroom means attributable to teacher characteristics.
Figure 1.1 Random Intercepts Model
Research Question 1: How consistently do three VAM predict teacher’s VAS?, the Simple Fixed Effects Model (SFEM, Tekwe, et al., 2004), the Covariate Adjustment Model (CAM, AIR, 2011) and the Hierarchical Linear Methods Model (HLMM, Tekwe, et al., 2004). Research question 1 was addressed in two steps. First, (VAS) were computed for each teacher using the three models, (the SFEM, the CAM and the HLMM) fitted with the same data. In the second step, the value-added scores were rank-ordered and placed into quintiles. These quintiles were then compared, using the intraclass correlation coefficients (ICCs) as a measure of agreement, to determine what extent teachers were assigned to the same quintile using three different models. Descriptions of the value-added calculations performed (statistical and SAS specifications provided) for each model are described below. A summary of the distinguishing features of the three models employed is presented in Table 1.1.

Model 1 (SFEM)
The parameterization used for the SFEM is:

\[ d_{ij} = \beta_{0is} + \sum_{l=1}^{39} \beta_{1ls} \cdot T_{l|t|j} + \varepsilon_{ij} \]

where
\[ d_{ij} = y_{ilst} - y_{ilst}, \]
\[ y_{ilst} = \text{the test score on the } s^{th} \text{ subject (PC or PV) at time } t \text{ for the } j^{th} \text{ student who had the } i^{th} \text{ teacher}, s = 1,2; \ t = 1,2; \ l = 1,2,...,40; \ j = 1,2,...,n_i, \]
\[ T_{l|t|j} = 1, 0, \text{ or } -1 \text{ as } l = k \text{ and } l \neq 40, l \neq k \text{ and } l \neq 40, \text{ or } l = 40, k = 1,2,...,39, \text{ respectively,} \]
and \( \varepsilon_{ij} \sim \text{NIID} \left(0, \sigma_{\varepsilon}^2\right) \) for each given \( s = 1,2, \)

The \( \beta_{1ls} \) coefficient is interpreted as the value-added in the \( s^{th} \) subject area by the \( k^{th} \) teacher.

Model 2 (HLMM)
The parameterization used for the HLMM is:

Student Level
\[ d_{ijs} = \beta_{0is} + \beta_{2s}\text{Student__Gender}_{ij} + \beta_{3s}\text{Student__Race}_{ij} + \varepsilon_{ijs} \]
**Teacher Level**

\[ \beta_{0is} = \gamma_{0s} + \gamma_{1s} \text{Teacher\_Gender}_{1i} + \gamma_{2s} \text{Teacher\_Race}_{2i} + \gamma_{3s} \text{Exp}_{3i} + \xi_{is} \]

**Combined Form**

\[ y_{ij2} = \gamma_{0s} + \gamma_{1s} \text{Teacher\_Gender}_{1i} + \gamma_{2s} \text{Teacher\_Race}_{2i} + \gamma_{3s} \text{Exp}_{3i} \]

\[ + \beta_{2s} \text{Student\_Gender}_{ij} + \beta_{3s} \text{Student\_Race}_{ij} + \xi_{is} \]

\[ b_{0is} \] is a random intercept associated with the \( i \)\(^{th} \) teacher and subject area \( s \).

\[ \beta_{1s}, \beta_{2s}, \beta_{3s} \text{, and } \beta_{4s} \] are the fixed effects of intake scores, Gender and Race on learning gain in subject area \( s \). The \( \xi_{is} \) is a random error \( (\xi_{is} \sim \text{NIID}(0, \sigma_{\xi}^2)) \) and \( \xi_{is} \sim \text{NIID}(0, \sigma_{\xi}^2) \) and are the value-added measures for the model. There is an underlying assumption that \( \xi_{is} \) and \( \xi_{is} \) are independent.

**Model 3 CAM**

The parameterization used for the CAM is:

**Student Level**

\[ y_{ij2} = \beta_{0is} + \beta_{1s} y_{ij1} + \beta_{2s} \text{Gender}_{ij} + \beta_{3s} \text{Race}_{ij} \]

**Teacher Level**

\[ \beta_{0is} = \gamma_{0s} + \gamma_{1s} \text{Gender}_{1i} + \gamma_{2s} \text{Race}_{2i} + \gamma_{3s} \text{Exp}_{3i} + \xi_{is} \]

**Combined Form**

\[ y_{ij2} = \gamma_{0s} + \gamma_{1s} \text{Gender}_{1i} + \gamma_{2s} \text{Race}_{2i} + \gamma_{3s} \text{Exp}_{3i} + \beta_{1s} y_{ij1} + \beta_{2s} \text{Gender}_{ij} + \beta_{3s} \text{Race}_{ij} + \xi_{is} \]

---

**Table 1.1 Models Employed and Summary of Their Characteristics**

<table>
<thead>
<tr>
<th>Model Identifier</th>
<th>Model Name</th>
<th>Dependent Variable</th>
<th>Intake Adjusted</th>
<th>School Effect</th>
<th>Student-Level Variables Included</th>
<th>Teacher-Level Variables Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>SFEM</td>
<td>Change Score</td>
<td>No</td>
<td>Fixed</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Model 2</td>
<td>HLMM</td>
<td>Change Score</td>
<td>Yes</td>
<td>Random</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Model 3</td>
<td>CAMM</td>
<td>Raw Post Score</td>
<td>Yes</td>
<td>Fixed</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
**Research Question 2:** To what extent do classroom characteristics and other variables predict teachers’ VAS? To address the second research question, the results of a regression analysis using aggregated data were analyzed. For the regression analysis, aggregated student and teacher level variables were included and results were analyzed to determine which factor(s) predicted teacher’s value-added scores. Student-level variables included gender, race, years from mean age, self-regulation (HSKT), and the standardized raw scores for PC and PV; Teacher-level variables included gender, race, teaching knowledge (TK), teaching experience and SSRS; and the school-level variable was SES.

**Research Question 3:** How well do the results of these three models predict teachers who teach in ways that are identified as high quality using other metrics (established measures of TQ)? The strength of the correlations was determined by establishing asymptotic confidence limits (Alf & Graf, 1999). This method allows for multiple correlations to be analyzed and for the degree to which the correlations differ to be understood. The model that yields the highest correlation with the established measures of teaching quality might be deemed a valid VAM.

More specifically, first, multi-level models were fit and run, then, predicted values were saved and correlations between the predicted and actual outcome values were squared to get a pseudo $r$-squared value. Next, the difference in these pseudo-$r$-squared values were found (between the model with predictors and an empty model) and confidence intervals were placed around the differences. Consistent with null value hypothesis testing, a confidence interval encompassing 0 indicates no difference in the models. If the CI does not include 0, then there is a difference in the model. Here, if the confidence interval around the difference in pseudo-$r$-squared values does not include 0, this (or these) variable(s) are said to be either positively or negatively predictive of quality teaching. The subsequent chapter reports the results from the analyses and details the findings for each research question.
CHAPTER FOUR
RESULTS

Results

The results of the analyses varied with regard to the model and its parameters. Of note, the following variables required the use of imputation techniques due to missing data: HSKT, and SSRS. The implications for imputing data are discussed in the following chapter. Furthermore, A total of 37 teachers were used out of a possible 40, due to incomplete or missing data for those three teachers. Thus, the answers to the three primary research questions posed are:

*Question 1:*

1. How consistently do three VAM using student reading and vocabulary outcomes predict individual teachers’ teaching quality (as deemed by VAS), the Simple Fixed Effects Model (SFEM, Tekwe et al., 2004), the Covariate Adjustment Model (AIR, 2011), and the Hierarchical Linear Methods Model (HLMM, Tekwe, et al., 2004).

First, each model was fitted with the same data. Models included only predictor variables (students’ standard scores on either PC or PV) for the purpose of answering this first research question. Next, all three models’ VAS (estimates produced via HLM analyses) were placed in Excel in descending order and were then sorted into quintiles to establish agreement (or lack thereof) across all three models. With a final total of 37 teachers, quintiles 1-4 contained 7 teachers each, and the 5th quintile contained 9 teachers. The first quintile represents those teachers whose VAS were high, and the fifth quintile represents VAS that were found to be low. Generally, the results from the three VAM did not consistently predict teaching quality (as determined by identifying the same teachers in each quintile). Percentages of agreement were determined by calculating how many teachers (by identical teacher IDs) were found in each quintile. Specifically, the results of these analyses are discussed below with regard to the percentage of agreement among and across quintiles for each of the three methods. In all model calculations, students’ standard scores were used, which allowed the researcher to control for students’ age. Therefore no gain from fall to spring is expected and indicates, that, on average, students in the classroom were making appropriate gains in either PC or PV for their age.
and grade. The mean teacher VAS for all models was -.082, or close to 0, as were all the VAS means (see Table 1.1, below). Hence, a positive VAS indicates that a teacher is more effective in improving his or her students test scores than the mean VAS. A negative VAS indicates that a teacher is less effective in improving test scores, in relation to the mean and hence the other teachers he (or she) is being compared to (students are showing lesser or negative standard score gains). This mean VAS of approximately zero is expected, as Tekwe et al. (2004) indicates the VAS are the random components of each model and each random component is assumed normally distributed with mean 0. The tables below provide descriptive information for each VAS type (by model) and highlight the differences in VAS by listing the mean VAS in each quintile as well as a grand mean per model. Of note is the mean VAS for the SFEM PC model was (M_{VAS}=0.307), possibly indicating a lack of fit for this model to the data.

<table>
<thead>
<tr>
<th>Quintile</th>
<th>Mean VAS Total</th>
<th>Mean VAS SFEM PC</th>
<th>Mean VAS SFEM PV</th>
<th>Mean VAS HLM PC</th>
<th>Mean VAS CAM PC</th>
<th>Mean VAS HLM PV</th>
<th>Mean VAS CAM PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (Higher VAS)</td>
<td>2.83</td>
<td>3.25</td>
<td>3.32</td>
<td>1.41</td>
<td>2.22</td>
<td>3.82</td>
<td>2.94</td>
</tr>
<tr>
<td>Q2</td>
<td>.847</td>
<td>.984</td>
<td>.954</td>
<td>.567</td>
<td>.918</td>
<td>.649</td>
<td>1.01</td>
</tr>
<tr>
<td>Q3</td>
<td>.031</td>
<td>.186</td>
<td>-.421</td>
<td>.242</td>
<td>.382</td>
<td>-.170</td>
<td>-.029</td>
</tr>
<tr>
<td>Q4</td>
<td>-.760</td>
<td>-.690</td>
<td>-.212</td>
<td>-.212</td>
<td>-.460</td>
<td>-1.08</td>
<td>-.904</td>
</tr>
<tr>
<td>Q5 (Lower VAS)</td>
<td>-1.93</td>
<td>-2.20</td>
<td>-2.29</td>
<td>-1.56</td>
<td>-2.38</td>
<td>-2.99</td>
<td>-2.35</td>
</tr>
<tr>
<td>Mean VAS</td>
<td>.307</td>
<td>.071</td>
<td>-.082</td>
<td>.136</td>
<td>.044</td>
<td>.134</td>
<td></td>
</tr>
</tbody>
</table>

The SFEM did not yield high percentages of agreement among identified teachers across quintiles, when compared with either the HLM or CAM models. Specifically, the
SFEM PC did not identify any teachers in the first quintile that either the HLM PC or CAM PC models identified (0% agreement). There was a 14% agreement between the SFEM PC and teachers identified by either the HLM PC or CAM PC methods in the second, third and fourth quintiles (that is, only 1 teacher was identified in both the SFEM PC and HLM PC models, and the same is true for the SFEM PC and CAM PC models, only 1 teacher was found in common in quintiles 2-4). In the fifth quintile, there was an 11% agreement among teachers identified by both the SFEM and the HLM PC (only 1 teacher was identified in both models), and a 22% agreement between those identified by the SFEM and the CAM PC (2 teachers were identified in both models). The table below provides a visual reference for the information detailed above.

<table>
<thead>
<tr>
<th>QUINTILE</th>
<th>SFEM PC/HLM PC</th>
<th>SFEM PC/CAM PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (Higher VAS)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Q2</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>Q3</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>Q4</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>Q5 (Lower VAS)</td>
<td>11%</td>
<td>22%</td>
</tr>
</tbody>
</table>

When examining the extent of agreement between the SFEM PV VAS and the HLM and CAM PV VAS, the results were wide-ranging. Specifically, the SFEM PV and the CAM PV models tended to identify the same teachers 13 to 43 percent of the time. See Table 1.4, below for percentages for each quintile.
Table 1.4 Percentage of Agreement for SFEM PV VAS and HLM PV and CAM

<table>
<thead>
<tr>
<th>QUINTILE</th>
<th>SFEM PV/HLM PV</th>
<th>SFEM PV/CAM PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (Higher VAS)</td>
<td>29%</td>
<td>43%</td>
</tr>
<tr>
<td>Q2</td>
<td>0%</td>
<td>29%</td>
</tr>
<tr>
<td>Q3</td>
<td>43%</td>
<td>29%</td>
</tr>
<tr>
<td>Q4</td>
<td>43%</td>
<td>29%</td>
</tr>
<tr>
<td>Q5 (Lower VAS)</td>
<td>0%</td>
<td>13%</td>
</tr>
</tbody>
</table>

The HLM and CAM models yielded highly similar results within quintiles among predictors. For instance, the HLM PC model identified 71% of the same teachers as the CAM PC model in the first quintile. The same is true for the HLM PV and the CAM PV models, as well, also yielding a 71% agreement in the first quintile. The results of remaining quintiles can be found in table 1.6, below.

Table 1.5 Percentage of Agreement for HLM PC/CAM PC and HLM PV/CAM

<table>
<thead>
<tr>
<th>QUINTILE</th>
<th>HLM PC/CAM PC</th>
<th>HLM PV/CAM PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (Higher VAS)</td>
<td>71%</td>
<td>71%</td>
</tr>
<tr>
<td>Q2</td>
<td>57%</td>
<td>29%</td>
</tr>
<tr>
<td>Q3</td>
<td>43%</td>
<td>29%</td>
</tr>
<tr>
<td>Q4</td>
<td>29%</td>
<td>57%</td>
</tr>
<tr>
<td>Q5 (Lower VAS)</td>
<td>78%</td>
<td>78%</td>
</tr>
</tbody>
</table>

A comparison of the SFEM models was also conducted to see how well the PC and PV models were able to predict one another by identifying the same teachers. Fairly low agreement was found for all quintiles, the SFEM models did not consistently identify the same teachers. See table 1.6, below for the breakdown of percentages per model and quintile.
Table 1.6 Percentage of Agreement between SFEM PC and SFEM PV

<table>
<thead>
<tr>
<th>QUINTILE</th>
<th>SFEM PC/SFEM PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (Higher VAS)</td>
<td>38%</td>
</tr>
<tr>
<td>Q2</td>
<td>13%</td>
</tr>
<tr>
<td>Q3</td>
<td>25%</td>
</tr>
<tr>
<td>Q4</td>
<td>13%</td>
</tr>
<tr>
<td>Q5 (Lower VAS)</td>
<td>25%</td>
</tr>
</tbody>
</table>

However, it is noteworthy that variability among the actual VAS did exist. For example, while teacher “A” was identified in the first quintile for both the HLM PC and CAM PC models, the HLM PC VAS for teacher “A” was 2.34, whereas the VAS for the CAM PC model was 1.47 for the same teacher.

Additionally, Intraclass Correlations (ICCs) were run to further establish how closely the teacher’s VAS agreed with one another. This statistical technique can be used when quantitative measurements are made on units that are organized into groups, here teachers’ VAS divided into quintiles. ICC results describe how strongly units in the same group resemble each other. Although ICCs are viewed as a type of correlation, unlike most other correlation measures, it operates on data structured as groups, rather than data structured as paired observations. Cut-offs have been established for ICC interpretation (Landis & Koch, 1977). Specifically, ICCs can be interpreted as follows: 0-0.2 indicates poor agreement; 0.3-0.4 indicates fair agreement; 0.5-0.6 indicates moderate agreement; 0.7-0.8 indicates strong agreement; and >0.8 indicates almost perfect agreement.

The single measures ICC (overall, for all teachers in all models) for the teachers in the study is .132, which, according to some researchers (Schrodt & Fliess, 1979; Landis et. al., 1977; and Portney et al., 2000) indicates a poor level of agreement, confirming the quintile analysis described above. However, when examining only PV VAS, there is an almost perfect agreement between HLM PV and CAM PV, with an ICC of .845. Also, there is near perfect agreement between HLM PC and CAM PC models, with an ICC of .898. Overall, there is little agreement between all six VAS (PC and PV for SFEM, HLM and CAM models). However, high agreement was found for both HLM PV/CAM PV models and HLM PC and CAM PC models. A summary of these results
can be found in Table 1.7, below. The implications for the extent of agreement among methods as well as the degree of variance in the actual VAS are considered in the discussion chapter that follows.

### Table 1.7 ICC Analyses

<table>
<thead>
<tr>
<th>Comparison</th>
<th>ICC (consistency)</th>
<th>ICC (absolute)</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFEM PC, SFEM PV, HLM PC, HLM PV, CAM PC, CAM PV</td>
<td>.132</td>
<td>.135</td>
<td>Poor</td>
</tr>
<tr>
<td>SFEM PC, HLM PC, CAM PC</td>
<td>.132</td>
<td>.135</td>
<td>Poor</td>
</tr>
<tr>
<td>SFEM PV, HLM PV, CAM PV</td>
<td>.373</td>
<td>.379</td>
<td>Fair</td>
</tr>
<tr>
<td>SFEM PC, SFEM PV</td>
<td>.109</td>
<td>.112</td>
<td>Poor</td>
</tr>
<tr>
<td>HLM PC, CAM PC</td>
<td>.898</td>
<td>.901</td>
<td>Near perfect</td>
</tr>
<tr>
<td>HLM PV, CAM PV</td>
<td>.845</td>
<td>.848</td>
<td>Near perfect</td>
</tr>
</tbody>
</table>

**Research Question 2:**

To what extent do classroom-wide student characteristics (i.e., classroom composition), and other variables not within the teachers’ control predict teachers’ value-added scores for Passage Comprehension?

For the purposes of answering this second research question, regression models were run with data aggregated to the teacher level. The following fall variables aggregated at the classroom level were added to the models to examine the effect of classroom-wide student characteristics: SES (Social Economic Status), Fall Class Self Regulation (HSKT), Class Fall PV, class proportion of Girls, class proportion of Caucasian students, Class Age from Mean (i.e., representing students who are younger or older than grade expectations), Teacher gender, Teacher Ethnicity, Teaching Experience
(Tyr), SSRS, and Teacher Knowledge (TK). Descriptive statistics are provided in Table 2.1, correlations in Table 2.2, and regression analysis in Table 2.3.

The SFEM model found both SES and Teacher Knowledge to be positively and significantly predictive of VAS. This finding indicates that teachers’ VAS tend to be higher when a greater percentage of students qualify for Free and Reduced Lunch (FARL). None of the other variables significantly predicted variability in the HLM nor CAM models’ VAS. Of note, (as is typically expected), the school with the lowest percentage of FARL had the highest incoming PC and PV scores. The school that had the highest percentage of students qualifying for FARL had the lowest scores (for both PC and PV).

Table 2.1 Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall HSKT</td>
<td>40</td>
<td>20.81</td>
<td>39.00</td>
<td>36.1227</td>
<td>3.75510</td>
</tr>
<tr>
<td>Fall PV</td>
<td>40</td>
<td>88.37</td>
<td>109.89</td>
<td>100.3701</td>
<td>5.03572</td>
</tr>
<tr>
<td>Student Gender</td>
<td>40</td>
<td>.09</td>
<td>.75</td>
<td>.4773</td>
<td>.15362</td>
</tr>
<tr>
<td>Student Ethnicity</td>
<td>40</td>
<td>.00</td>
<td>.98</td>
<td>.4596</td>
<td>.30697</td>
</tr>
<tr>
<td>Years from Mean Age</td>
<td>40</td>
<td>-.42</td>
<td>.46</td>
<td>.1149</td>
<td>.22024</td>
</tr>
<tr>
<td>Teacher Gender</td>
<td>37</td>
<td>.00</td>
<td>1.00</td>
<td>.9730</td>
<td>.16440</td>
</tr>
<tr>
<td>Teacher Ethnicity</td>
<td>37</td>
<td>.00</td>
<td>1.00</td>
<td>.7838</td>
<td>.41734</td>
</tr>
<tr>
<td>SSRS</td>
<td>40</td>
<td>187.60</td>
<td>218.23</td>
<td>199.9745</td>
<td>6.60473</td>
</tr>
<tr>
<td>Teacher Knowledge</td>
<td>39</td>
<td>6.45</td>
<td>43.06</td>
<td>30.2288</td>
<td>9.02242</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.2 Correlation Coefficients

<table>
<thead>
<tr>
<th></th>
<th>SFEM_PC</th>
<th>HLM_CAT_PC</th>
<th>CA_MCAT_PC</th>
<th>Fall_HSK_T_Total_mean</th>
<th>Fall_P_V_SS_mean</th>
<th>Yrfrom_MeanAge_mean</th>
<th>TK</th>
<th>SSRS2_mean</th>
<th>T_Years_mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFEM_PC</td>
<td>1</td>
<td>.964**</td>
<td>.956*</td>
<td>.014</td>
<td>-.229</td>
<td>-.228</td>
<td>.341</td>
<td>-.217</td>
<td>-.169</td>
</tr>
<tr>
<td>HLM_CAT_PC</td>
<td>.964**</td>
<td>1</td>
<td>.940**</td>
<td>.050</td>
<td>-.069</td>
<td>-.145</td>
<td>.189</td>
<td>-.257</td>
<td>-.083</td>
</tr>
<tr>
<td>CAM_CAT_PC</td>
<td>.956**</td>
<td>.940**</td>
<td>1</td>
<td>.036</td>
<td>.106</td>
<td>-.166</td>
<td>.283</td>
<td>-.156</td>
<td>.067</td>
</tr>
<tr>
<td>Fall_HSK</td>
<td>.014</td>
<td>.050</td>
<td>.036</td>
<td>1</td>
<td>.382*</td>
<td>-.210</td>
<td>.114</td>
<td>-.063</td>
<td>.017</td>
</tr>
<tr>
<td>Yrfrom_MeanAge</td>
<td>-.229</td>
<td>-.069</td>
<td>.034</td>
<td>.329</td>
<td>1</td>
<td>.154</td>
<td>-.104</td>
<td>.233</td>
<td>.134</td>
</tr>
<tr>
<td>SSRS2_mean</td>
<td>-.233</td>
<td>-.299</td>
<td>-.249</td>
<td>-.204</td>
<td>.154</td>
<td>1</td>
<td>-.246</td>
<td>.026</td>
<td>-.004</td>
</tr>
<tr>
<td>Teacher_Knowl</td>
<td>.341</td>
<td>.189</td>
<td>.283</td>
<td>.114</td>
<td>-.104</td>
<td>-.246</td>
<td>1</td>
<td>.054</td>
<td>-.109</td>
</tr>
<tr>
<td>T_Years_mean</td>
<td>-.217</td>
<td>-.257</td>
<td>-.156</td>
<td>-.063</td>
<td>.233</td>
<td>.026</td>
<td>.054</td>
<td>1</td>
<td>.171</td>
</tr>
</tbody>
</table>

** Note: Correlation coefficients are significant at the .05 level.
Table 2.3 Regression Results

<table>
<thead>
<tr>
<th></th>
<th>SFEM-PC</th>
<th></th>
<th>HLM-PC</th>
<th></th>
<th>CAM-PC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>Sig.</td>
<td>B</td>
<td>SE</td>
<td>Sig.</td>
</tr>
<tr>
<td>Intercept (Constant)</td>
<td>-.485</td>
<td>15.399</td>
<td>.975</td>
<td>13.190</td>
<td>19.466</td>
<td>.506</td>
</tr>
<tr>
<td>School SES</td>
<td>.067</td>
<td>.025</td>
<td>.014*</td>
<td>.053</td>
<td>.027</td>
<td>.066</td>
</tr>
<tr>
<td>Class Self-Regulation</td>
<td>-.011</td>
<td>.040</td>
<td>.778</td>
<td>-.026</td>
<td>.048</td>
<td>.589</td>
</tr>
<tr>
<td>Class PV</td>
<td>-.025</td>
<td>.100</td>
<td>.805</td>
<td>-.102</td>
<td>.103</td>
<td>.331</td>
</tr>
<tr>
<td>Class Ethnicity</td>
<td>4.945</td>
<td>2.449</td>
<td>.053</td>
<td>4.964</td>
<td>2.494</td>
<td>.060</td>
</tr>
<tr>
<td>Teacher Ethnicity</td>
<td>-1.452</td>
<td>.790</td>
<td>.077</td>
<td>-.419</td>
<td>.774</td>
<td>.594</td>
</tr>
<tr>
<td>Class Age from Mean</td>
<td>-.756</td>
<td>1.600</td>
<td>.640</td>
<td>-1.764</td>
<td>1.871</td>
<td>.357</td>
</tr>
<tr>
<td>Teacher Knowledge</td>
<td>.099</td>
<td>.040</td>
<td>.019*</td>
<td>.004</td>
<td>.055</td>
<td>.937</td>
</tr>
<tr>
<td>Teaching Experience</td>
<td>-.013</td>
<td>.031</td>
<td>.669</td>
<td>.063</td>
<td>.032</td>
<td>.067</td>
</tr>
</tbody>
</table>

Note. No predictors significantly explained variability in VAS (p ≤ .05 level).

Of note, the CAM models utilize a post-test score (Spring SS) as the outcome, with the pre-test score as a covariate whereas the HLM models utilize a gain score as the outcome with no pre-score as a covariate. Additionally, passage comprehension was selected as a child-level outcome in this study (for computing the VAS in RQ1) as such assessments are highly correlated with state-mandated tests that are typically used in VAM calculations. However, picture vocabulary was also chosen for use in this study (for regression analysis in RQ2) as research indicates that early vocabulary knowledge
predicts future reading success (Senechal et al., 2006). Particularly in second grade, students are in the process of acquiring expressive vocabulary skills that are essential for enabling successful reading comprehension. Implications for these findings are discussed in chapter five.

Research Question 3:
How well do the results of these three VAS models, using PC as the student outcome, concur with characteristics of teachers, specifically those who have more versus less specialized knowledge about reading (TK) and higher quality teaching practices based on observations of classroom practices? To answer this third research question, Passage Comprehension (PC) and Picture Vocabulary (PV) were both used as student-level outcome measures.

For each of the models, teacher knowledge (TK), teaching quality (Total Quality, TQ), and a combined predictor, with TK and TQ were run in the method proposed by Alf and Graf (1999), a multiple-correlation technique. Several steps were taken to arrive at the findings described below. First, multi-level models were run, and the VAS were saved. Then, empty and full models were established (empty model contains no predictor variables, essentially, the SFEM model), and the full model contained the following predictor variables: teacher knowledge (TK), teaching quality (TQ), and a combination of the two. Next, correlations were run on the VAS. Then, those values were squared to get the pseudo-$r^2$ values. Last, the Alf and Graf method of confidence interval estimation around multiple correlations was employed, which yielded the results found in table 3.1, below.

In this analysis, predictors were considered significantly different from 0 if the difference in the pseudo-$r^2$ value between the model with that predictor (TK, TQ, or the combination of the two) and the empty model were significantly different from zero. For example, TK as an external predictor (in the SFEM model) provides an increase in percentage of variance in the VAS explained of 11.6%, as indicated by the difference in pseudo-$r^2$ values between the model with TK and the model without (95% CI: -.309 to -.077). None of the models were associated with the teaching quality measures (TK and TQ). That is, none of the variables (TK, TQ) were found to be consistently positive or significant. This indicates the models with TK and TQ are equal to an empty (null)
model when predicting teacher VAS (equal to the models without these variables).
Unfortunately, the models that included additional variables (TK and TQ) did not provide
any additional explanatory power and the empty (null) model performed just as well in
terms of predicting VAS.
<table>
<thead>
<tr>
<th>Model</th>
<th>External Predictors</th>
<th>N</th>
<th>Asymptotic Variance Components</th>
<th>Difference in $r_0A$ and $r_0B$</th>
<th>SE(Difference)</th>
<th>95% lower</th>
<th>95% upper</th>
<th>Significant predictor?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Teacher knowledge</td>
<td>37</td>
<td>1.081</td>
<td>0.009</td>
<td>-6.44</td>
<td>0.098</td>
<td>-0.309</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Total Quality</td>
<td>37</td>
<td>1.081</td>
<td>0.001</td>
<td>-2.16</td>
<td>0.010</td>
<td>0.032</td>
<td>-0.074</td>
</tr>
<tr>
<td>SFEM</td>
<td>Teacher knowledge and total quality</td>
<td>37</td>
<td>1.081</td>
<td>0.011</td>
<td>-6.77</td>
<td>0.010</td>
<td>0.137</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>Teacher knowledge</td>
<td>37</td>
<td>0</td>
<td>0.003</td>
<td>0</td>
<td>0.003</td>
<td>0.035</td>
<td>0.059</td>
</tr>
<tr>
<td>HLM</td>
<td>Total Quality</td>
<td>37</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
<td>0.001</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>Teacher knowledge and total quality</td>
<td>37</td>
<td>0</td>
<td>0.004</td>
<td>0</td>
<td>0.004</td>
<td>0.041</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>Teacher Knowledge</td>
<td>37</td>
<td>0</td>
<td>0.007</td>
<td>0</td>
<td>0.007</td>
<td>0.080</td>
<td>0.085</td>
</tr>
<tr>
<td>CAM</td>
<td>Total Quality</td>
<td>37</td>
<td>0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
<td>0.003</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>TK &amp; TQ</td>
<td>37</td>
<td>0</td>
<td>0.007</td>
<td>0</td>
<td>0.007</td>
<td>0.080</td>
<td>0.085</td>
</tr>
</tbody>
</table>
Table 2.5 Multiple Correlation Analysis Result for PV Models

<table>
<thead>
<tr>
<th>Model</th>
<th>External Predictors</th>
<th>N</th>
<th>Asymptotic Variance Components</th>
<th>Difference in r0A and r0B</th>
<th>SE(Difference)</th>
<th>95% lower</th>
<th>95% upper</th>
<th>Significant predictor?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFEM</td>
<td></td>
<td>37</td>
<td>9.729</td>
<td>0.001</td>
<td>-6.286</td>
<td>0.000</td>
<td>-0.009</td>
<td>0.031</td>
</tr>
<tr>
<td>PV</td>
<td>Teacher knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total Fidelity</td>
<td>37</td>
<td>9.729</td>
<td>0.005</td>
<td>-0.000</td>
<td>0.004</td>
<td>-0.052</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>Teacher knowledge and total fidelity</td>
<td>37</td>
<td>9.729</td>
<td>0.005</td>
<td>-0.000</td>
<td>0.005</td>
<td>-0.057</td>
<td>0.073</td>
</tr>
<tr>
<td>HLM</td>
<td></td>
<td>37</td>
<td>0.000</td>
<td>0.003</td>
<td>-0.00</td>
<td>0.002</td>
<td>-0.029</td>
<td>0.045</td>
</tr>
<tr>
<td>PV</td>
<td>Teacher knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total Fidelity</td>
<td>37</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.00</td>
<td>0.001</td>
<td>-0.017</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>Teacher knowledge and total fidelity</td>
<td>37</td>
<td>0.000</td>
<td>0.004</td>
<td>-0.00</td>
<td>0.002</td>
<td>-0.041</td>
<td>0.054</td>
</tr>
<tr>
<td>CAM</td>
<td></td>
<td>37</td>
<td>0</td>
<td>0.004</td>
<td>0</td>
<td>0.004</td>
<td>-0.041</td>
<td>0.063</td>
</tr>
<tr>
<td>PV</td>
<td>Teacher knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total Fidelity</td>
<td>37</td>
<td>0</td>
<td>0.007</td>
<td>0</td>
<td>0.007</td>
<td>-0.083</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>Teacher knowledge and total fidelity</td>
<td>37</td>
<td>0</td>
<td>0.009</td>
<td>0</td>
<td>0.009</td>
<td>-0.111</td>
<td>0.097</td>
</tr>
</tbody>
</table>
CHAPTER FIVE

DISCUSSION

In 2012, value-added measures have become one of the most popular methods for evaluating teaching quality. The practice of using VAM to measure teaching quality is increasingly being used in school districts across the country, as many states have adopted legislation dictating their use. Florida is among at least twenty-three states that are using student scores on standardized exams to help evaluate teachers (Miami Herald, November 6th, 2011) and by 2014, it will become mandatory for even more states to do so under new legislation (SB 736). In Florida, the stakes are high. Top-performing teachers can receive permanent salary increases, while those at the bottom for two consecutive years risk losing their jobs entirely. Although evaluating the impact of teachers on their students may seem somewhat intangible and is no doubt a challenging task, it is nonetheless essential to determine which teachers are consistently producing higher student achievement gains, as these teachers hold the key to our understanding of high quality instruction. No two students are exactly alike, as we are all unique individuals, however there are teachers that outperform their peers every year. These teachers are the ones that researchers are most interested in studying to understand what they do differently to reach such results.

It is widely accepted among policy-makers and researchers that value-added assessments of teacher performance are preferred to assessments based on status-scores alone. However, what still needs to be determined is which method should be employed for such calculations. Given the plethora of options for calculating value-added scores, this is no easy task. Conclusions from this study and their implications are discussed below. Several general conclusions can be drawn from the results of the current study concerning value-added assessments of teaching quality as well as several specific inferences per research question.

Summary of Findings

Consistency of VAM predicting VAS. The three VAM employed in this study (SFEM, HLM, CAM) did not consistently identify the same teacher. This finding indicates the instability of these value-added methods and highlights the importance of model choice and variable selection for inclusion within these models. Importantly, very
little agreement was found when the SFEM models were compared to the CAM and HLMM models. Additionally, there was very little agreement between all six VAS. These findings indicate that not only does the model type make a difference in identifying effective teaching but so does the standardized assessments used to compute teacher VAS. However, there was very high agreement (near-perfect) between the HLM PV and CAM PV models as well as between the HLM PC and CAM PC models.

The findings from this first research question support the following conclusions:
1. Different VAM identify different teachers, as demonstrated by this study and the calculation of the SFEM, CAM, and HLM models for teachers’ VAS. 2. The model employed for calculating teachers’ VAS really matters. The importance of model selection cannot be overstated, as each model, particularly the SFEM as compared with the CAM and HLM models, resulted in different VAS for the same teachers whereas the CAM and HLM models showed almost perfect agreement as long as the same standardized assessment was used. For example, the ICC comparing the HLM VAS calculated using PV and PC was very low. This has implications for both policy and practice. From a policy standpoint, a consensus needs to be reached regarding a standard go-to method for VAS calculation. These models continue to be implemented in school districts all across America. However, without agreed-upon models and variables (i.e., standardization for VAM), little value will be found in their outcomes. From a practical perspective, the teachers who are being judged by these scores (and in some situations, teachers are judged via reading scores even if they don’t teach reading) may receive an unfair assessment of their performance (as these VAS seem to be fairly arbitrary). The findings from this study suggest that researchers decide which type of model to use and a standard set of variables to include in the calculation of VAS or it will render these calculations useless.

Variables Predicting VAS. When analyzing the results of the second research question, the variables responsible for predicting the VAS in each model were considered, including, SES, teacher/student gender and ethnicity, teacher knowledge and years of experience, teacher’s ability to judge their students’ academic skills (SSRS), classroom composition (HSKT), and years from mean age (Table 2.2). The SFEM found both SES (percentage of students qualifying for FARL) and teacher knowledge (TK) to
be positively significant predictors of VAS. At least in this district, teachers at higher poverty schools were more likely to have higher VAS than teachers at more affluent scores. Counterintuitive, this is a potentially important finding that belies the extant research that students attending higher poverty schools generally show less effective instruction and weaker gains in achievement (Connor, Jakobsons, Crowe, & Meadows, 2009). There are several potential reasons for this finding. The school district may have intentionally placed its most effective teachers where they were most needed based on student achievement, there may have been regression to the mean, or there may have been more room for these students to achieve. Additional research is needed to understand and explore this finding.

Additionally, teachers who demonstrated greater levels of specialized knowledge on a test of reading instruction knowledge were more likely to have higher VAS than teachers with lower scores. This finding is encouraging because teacher knowledge can be increased through targeted training and effective professional development (Fishman et al., 2003). Teaching quality may also be enhanced if teachers are provided the opportunity to participate in active learning that focuses on improving content knowledge of the profession (Garet et. al., 2001). However, teacher/student gender and ethnicity, teacher knowledge and years of experience, SSRS, HSKT, and years from mean age did not significantly predict VAS from the HLMM or CAM models. It may be that the sample size of only 37 teachers may not provide enough power to detect significant associations for any of the variables included in this study. Thus there may be associations that are important but for which there was inadequate power to find potentially educationally important effects. At the same time, there were no noticeable trends found for any of the models, resulting in a mix of positive and negative coefficients.

In general and for this sample (only 1 male teacher), female teachers had higher value added scores than did male teachers. This finding is interesting to explore for implications. For instance, if an education policy were to be based on this finding, it might recommend the removal of male teachers from the profession, which is simply preposterous. This finding underscores the importance of carefully analyzing the
variables included in a model as well as scrutinizing of the practical results those models yield.

*Other Measures of Teaching Quality.* The third research question investigated whether VAS computed using three different PC and PV VAM were associated with other measures of literacy teaching quality. The variables typically expected to be associated with higher VAS, Teacher Knowledge (TK) and overall Classroom Learning Environment Quality (TQ) were not associated with higher VAS for either the PC or PV models. This finding is thought provoking and warrants further research to investigate the implications if these results are replicated. This means that higher VAS scores are not associated with classroom and teaching characteristics that are generally found, in the literature, to be associated with higher quality teaching. This finding is discouraging, and again points to the inability of these models to consistently detect quality teachers, as both teacher knowledge (TK) and TQ are well-regarded in the literature for their ability to identify high quality teaching.

**Limitations**

It is important to keep in mind that the results from this study are specific to a small sample of second-grade teachers and their students’ standardized reading achievement scores (WJIII PV and PC), thus limiting generalizability. It is not clear whether conclusions drawn from this study of teachers in a single Florida school district will generalize to teachers of other districts. This study might be replicated with a larger number of teachers and their students to understand the answers to the questions posed here more clearly.

The value-added scores calculated in this study are specific to 37 second-grade teachers and their students’ reading and vocabulary scores. It is not known how well the results of literacy-based standardized scores are transferable to other subjects or other grade levels, particularly outside of the elementary realm. Given the heavy reliance on reading scores for assessing teaching quality (even for teachers who do not teach reading), it is necessary to keep in mind that reading scores may not be appropriate for assessing teachers who teach in subject areas, such as history, or science (Maerten et al., 2010). In order for VAS to have the potential to provide valuable information regarding
teaching quality, the standardized examinations used in such calculations must directly assess the construct teachers are responsible for teaching their students.

As with any calculation of value-added methods using student-level data to determine value-added scores, student achievement test scores were not intended for such measurement (evaluating teaching quality). In general, the teacher sample size was small and so the aggregated data sample may not have provided the statistical power to detect significant differences and associations where they might actually have existed.

Throughout this discussion, the models are said to produce value-added measures. This implicitly refers to the model itself coupled with the estimation procedure calculated by SAS, HLM or PASW to estimate the model’s parameters. The model and estimation procedure together produce value-added measures and hence, the value-added scores used here to evaluate teaching quality. Additional research is needed to determine whether the SFEM, HLMM, and CAM models would produce notably different results when more than two years of data are analyzed.

Conclusion

The results of this study highlight clear concerns with VAS albeit with a small sample of teachers. At the same time, the nation’s reliance on VAM continues to increase; it appears as though this method for evaluating teaching quality is here to stay. Given the concerns that (1) different VAM with different outcomes do not identify the same teachers as highly effective; and (2) that teacher VAS using reading as the outcome are not associated with identified characteristics of effective teaching. Researchers and educators might need to establish clear guidelines for estimating teaching quality via these VAM formulas. In order to facilitate this, it may be necessary to employ several strategies for arriving at teaching quality decisions, as this study has demonstrated. Here, the use of HLM analyses and regression analyses in conjunction with multiple-correlation analyses provided a conflicting account of which teachers might actually be effective in improving student outcomes and made it more difficult for conclusions about teaching quality to be reached. For example, none of the variables in either the regression analyses or the multiple correlation analysis significantly predicted the VAS (research questions 2 and 3). However, when the VAS were examined closely in the first research question, there was high agreement between the HLMM and CAM models, but not when
the SFEM models were compared to either HLMM or CAM. Hence, a similar combination of methods might be needed to adequately employ VAM, especially in instances where employment decisions are based on VAS. For example, bonus and firing decisions might be made more appropriately using multiple measures in addition to VAS.

It is concerning that even with established standardized measures for reading assessment (WJIII PC and PV) in the elementary domain (here, second grade) that such discrepancies exist between the VAS and the teachers these VAM identify. For many subjects, there are no established, agreed-upon assessments, such as science, history, art, music and others, and in subjects where standardized assessments are plentiful, the research is not yet clear how well current high-stakes assessments measure students’ knowledge of science (Brickhouse, 2006). Additionally, standardized measures are few and far between for grades beyond elementary (5th-12th). If the results from these three VAM (SFEM vs. CAM/HLMM) yield such inconsistent results, (considering the use of established, agreed-upon measures) this causes serious concerns for the explicit wide spread use of these methods for the purposes of assessing teaching quality, particularly outside of the K-5th arena. The findings from this study underscore the need to develop VAM that yield comparable results within a small margin of error (by identifying the same teachers with consistency).

Although the results from this study are discouraging with regard to implementing VAM for the purposes of evaluating teaching quality, policy makers and educational administrators will always want a way to quantify teaching quality, and therefore this researcher recognizes the potential value in these methods. Additionally, if reliance on VAM continues, efforts need to be made towards the development of adequate standardized measures that accurately assess students’ knowledge in subjects outside of the literacy realm. According to some, measures that serve as sufficient proxies for actual student learning in subjects like science do not yet exist (Brickhouse, 2006). Of course, this very cautious recommendation of VAM is not without a few important caveats. Most notably, a consensus must be reached regarding VAM (type of method, and variable inclusion), or comparisons cannot be made from district to district and from state to state, thus rendering these methods impractical. VAM are not perfect, but with careful consideration of the data as well as models employed (and perhaps the inclusion
of other methods of evaluating teaching quality), VAM have the potential to become a viable method available for evaluating teaching quality. However, there is still much work to be done to develop VAM that consistently yield comparable results, so that they may be employed with confidence in the future.
APPENDIX A

TEACHER KNOWLEDGE SURVEY (TKS)
Multiple Choice. Please circle the letter of the best answer on the line. As stated earlier, you are not expected to know the answers to all of the questions below. For each question, there is a response that states: I don’t know. Additionally, there is a space for you to identify to whom or what you would refer to for assistance (i.e teacher’s manual, reading coach, other teacher, speech language pathologist, etc.). Please choose both a response to the question and fill in this response. If you do not know where you would go for assistance, or would not seek a reference please mark the line with I don’t KNOW or NONE respectively. Please answer the questions in the order in which they appear. Thank you for your participation!

1. A fellow teacher approaches you and says that she is teaching a lesson on schwa, and asks for examples of words that have a schwa sound in them. You tell her that a good example of a word with a schwa is:
   (a) resume  (d) about
   (b) bread  (e) flirt
   (c) look  (f) I don’t know.
   I would seek the following as a reference: _________________________

2. You are focusing a lesson on short vowel sounds. Which word might you use as part of your lesson?
   (a) treat  (d) paw
   (b) start  (e) father
   (c) slip  (f) I don’t know.
   I would seek the following as a reference: _________________________

3. A reading coach asks you how you would explain phonemes to your 5th grade class. The reading coach wants to be sure that your understanding of a phoneme is correct. You tell her that a phoneme refers to:
   (a) a single letter   (c) a single unit of meaning
   (b) a single speech sound  (d) a grapheme
   (e) I don’t know.
   I would seek the following as a reference: _________________________

4. You and several of the teachers in your school have just attended a workshop on decoding that provided definitions for phonemes, graphemes, syllables, and morphemes. Some of the other teachers are confused by one of the definitions from the workshop and ask you for help. They want to know what a pronounceable group of letters containing a vowel sound is. You tell them this is a:
   (a) phoneme   (c) syllable
   (b) grapheme   (d) morpheme
   (e) I don’t know.
   I would seek the following as a reference: _________________________

5. You are playing nonsense word bingo in your class. When you are calling out nonsense words, you come across the nonsense word tife. You should pronounce the letter i in this word probably sounding like the i in:
6. You are giving a professional development workshop. If you were explaining a combination of two or three consonants pronounced so that each letter keeps its own identity, you would likely be explaining a:
   (a) silent consonant  (c) diphthong
   (b) consonant digraph (d) consonant blend
   (e) I don’t know.
I would seek that information from: _________________________

7. You are reviewing a textbook for possible use at your school. Which of the following words could you use for the schwa lesson because it contains a schwas sound?
   (a) cotton  (d) preview
   (b) phoneme  (e) grouping
   (c) stopping  (f) I don’t know.
I would seek that information from: _________________________

8. One of your students needs help on her reading work, which focuses on diphthongs. You tell her, for an example, that a diphthong is found in the word:
   (a) coat  (d) sing
   (b) boy  (e) been
   (c) battle  (f) I don’t know.
I would refer the student to: _________________________

9. Complete this sentence. A voiced consonant digraph is in the word:
   (a) think  (d) the
   (b) ship  (e) photo
   (c) whip  (f) I don’t know.
I would seek that information from: _________________________

10. You are giving a lesson talking about two combined letters that represent one single speech sound. You are referring to a:
    (a) schwa  (d) digraph
    (b) consonant blend (e) diphthong
    (c) phonetic    (f) I don’t know.
I would seek that information from: _________________________

11. How many speech sounds are in the word eight?
    (a) two  (c) four
    (b) three  (d) five
    (e) I don’t know.
I would seek that information from: _________________________
12. How many speech sounds are in the word *box*?
   (a) one   (c) three
   (b) two   (d) four
   (e) I don’t know.
   I would seek that information from: __________________

13. How many speech sounds are in the word *grass*?
   (a) two   (c) four
   (b) three  (d) five
   (e) I don’t know.
   I would seek that information from: __________________

14. At a grade level meeting, teachers are complaining that their students are confusing the sounds /b/ and /p/ or /f/ and /v/. You tell them that this may be happening because:
   (a) Students are visually scanning the letters in a way that letters are misperceived.
   (b) The students can’t remember the letter sounds so they are randomly guessing.
   (c) The speech sounds within each pair are produced in the same place and in the same way, but one is voiced and the other is not.
   (d) The speech sounds within each pair are both voiced and produced in the back of the mouth.
   (e) I don’t know.
   I would refer the teachers to: _________________________

15. You say to your students, “I am going to say a word and then I want you to break the word apart. Tell me each of the sounds in the word *dog.*” What type of task are you asking for?
   (a) blending  (c) segmentation
   (b) rhyming   (d) deletion
   (e) I don’t know.
   I would seek that information from: _________________________

16. You say to your students, “I am going to say some sounds that will make one word when you put them together. What does /sh/ /oe/ say?” What type of task are you asking for?
   (a) blending  (c) segmentation
   (b) rhyming   (d) manipulation
   (e) I don’t know.
   I would seek that information from: _________________________

17. Mark the statement that is FALSE and you would expect the reading coach NOT to say this to you or your fellow teachers.
   (a) Phonological awareness is a precursor to phonics.
   (b) Phonological awareness is an oral language activity.
   (c) Phonological awareness is a method of reading instruction that begins with individual letters and sounds.
(d) Many children acquire phonological awareness from language activities and reading.
(e) I don’t know.

I would seek that information from: ____________________________

18. A reading method that focuses on teaching the application of speech sounds to letters is called:
   (a) phonics     (d) phonetics
   (b) phonemics   (e) either (a) or (d)
   (c) orthography (f) I don’t know.

I would seek that information from: ____________________________

19. Your curriculum specialist is developing a spelling curriculum. When you review this curriculum, the rule you expect for using a *ck* in spelling would be:
   (a) to use *ck* when the vowel sound is a diphthong  (c) to use *ck* when the vowel sound is long
   (b) to use *ck* when the vowel sound is short   (d) any of the above
   (e) I don’t know.

I would seek that information from: ____________________________

20. Count the number of *syllables* in the word *unbelievable*.
   (a) four  (c) six
   (b) five   (d) seven
   (e) I don’t know.

I would seek that information from: ____________________________

21. Count the number of *syllables* in the word *pies*.
   (a) one  (c) three
   (b) two   (d) four
   (e) I don’t know.

I would seek that information from: ____________________________

The next two items involve saying a word and then reversing the order of the sounds. For example, the word *back* would be *cab*.

22. If you say the word, and then reverse the order of the sounds, *ice* would be:
   (a) easy  (c) size
   (b) sea    (d) sigh
   (e) I don’t know.

I would seek that information from: ____________________________

23. If you say the word, and then reverse the order of the sounds, *enough* would be:
   (a) fun    (c) funny
   (b) phone  (d) one
   (e) I don’t know.

I would seek that information from: ____________________________
24. In your classroom, students are identifying sounds. You are sitting in with a group during their center time. The child’s task is to identify the second sound in the word *queen*. The child would be correct by responding:

(a) u   (c) k
(b) long e   (d) w
(e) I don’t know.

I would refer the child to: _________________________

25. In your classroom, students are identifying sounds. You are sitting in with a group during their center time. The child’s task is to identify the third speech sound in the word *wretch*. The child would be correct by responding:

(a) /ch/   (c) /t/
(b) /e/   (d) /r/
(e) I don’t know.

I would refer the child to: _________________________

26. A fellow teacher wants to confirm that she is correct. She says, “In the word *crouch*, the *cr-* part is called the ________.” You respond by saying:

(a) rhyme   (d) morpheme
(b) initial phoneme   (e) onset
(c) rime   (f) I don’t know.

I would refer the teacher to: _________________________

27. In language, a single unit of meaning is called a

(a) grapheme   (d) morpheme
(b) syllable   (e) phoneme
(c) rime   (e) I don’t know.

I would seek that information from: ____________________

28. Count the number of *syllables* in the word *walked*.

(a) one   (c) three
(b) two   (d) four
(e) I don’t know.

I would seek that information from: _________________________

29. You say to your students, “The word is *taught*. What word would you have if you said *taught* without the /t/ sound?” What type of task are you asking for?

(a) rhyming   (c) elision
(b) blending   (d) none of the above
(e) I don’t know.

I would seek that information from: _________________________

30. A fellow teacher wants to confirm that she is correct. She says, “In the word *plan*, the –*an* part is called the ________.” You respond by saying:

(a) rhyme   (d) morpheme
(b) final phoneme  (e) onset 
(c) rime  (f) I don’t know.

I would seek that information from: ____________________

31. At a faculty meeting, some teachers want to know more information about skilled readers versus developing readers. You tell them that for skilled readers, listening and reading comprehension are usually about equal. For developing readers in K-3, it is true that:

(a) Reading comprehension is better than listening comprehension.
(b) Listening comprehension is better than reading comprehension.
(c) Reading and listening comprehension are comparable, about the same.
(d) There is no systematic relationship between reading comprehension and listening comprehension.
(e) I don’t know.

I would seek that information from: ____________________

32. How many morphemes are in the word gardener?

(a) one  (c) three
(b) two  (d) four
(e) I don’t know.

I would seek that information from: ____________________

33. How many morphemes are in the word unbelievable?

(a) one  (c) three
(b) two  (d) four
(e) I don’t know.

I would seek that information from: ____________________

34. How many morphemes are in the word pies?

(a) zero  (c) two
(b) one  (d) three
(e) I don’t know.

I would seek that information from: ____________________
**Short Answer. Please answer to the best of your ability.**

35. You are working on syllable types in your classroom, and even though you will not cover all of them, you want to hang a poster of all six syllable types in your classroom for the students to see. List the **six syllable types** and an example of each (e.g., a single-syllable word exemplifying the particular syllable type, a multi-syllable word with the specified syllable type circled). As an example, the first has been listed for you (with any one of the labels considered correct); if you are able, please provide an example of this syllable type before moving onto the others.

1) Closed syllable, CVC, or VC ________________________________

2) ________________________________

3) ________________________________

4) ________________________________

5) ________________________________

6) ________________________________
APPENDIX B

TEACHER BACKGROUND QUESTIONNAIRE
Name:  

School:  

1  Gender:  

M  F  

Race:  

o White  
o African-American  
o Latino/Hispanic  
o Asian/Pacific Islander  
o Other (please specify):  

7  A  

8  Total number of years of teaching experience:  

9  Total number of years spent teaching at current school:  

10  Length of literacy block:  

   minutes  

11  Are you expected to turn in your lesson plans to an administrator?  

Y  N  TBQ_12A  1=yes  

12  How many adults (co-teachers, volunteers, etc., including yourself) are generally present in your classroom during the literacy block?  

   

13  Degrees held (please check all that apply and specify your major/field/concentration for each):  

o B.A./B.S.  

   Major:  

15  A  

70
<table>
<thead>
<tr>
<th>Degree</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.A./M.S.</td>
<td>Field:</td>
</tr>
<tr>
<td>M.Ed./M.A.T.</td>
<td>Field:</td>
</tr>
<tr>
<td>C.A.G.S.</td>
<td>Field:</td>
</tr>
<tr>
<td>Ph.D.</td>
<td>Field:</td>
</tr>
<tr>
<td>Ed.D.</td>
<td>Field:</td>
</tr>
<tr>
<td>Other</td>
<td>Specify degree/field:</td>
</tr>
</tbody>
</table>

**Certifications held (check all that apply):**

<table>
<thead>
<tr>
<th>Certification</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary Education (grades K-6)</td>
<td>28A</td>
</tr>
<tr>
<td>Prekindergarten/Primary Education (age 3-grade 3)</td>
<td>29A</td>
</tr>
<tr>
<td>Preschool Education (birth through age 4)</td>
<td>30A</td>
</tr>
<tr>
<td>Middle Grades English (grades 5-9)</td>
<td>33A</td>
</tr>
<tr>
<td>English (grades 6-12)</td>
<td>32A</td>
</tr>
<tr>
<td>Middle Grades Integrated Curriculum (grades 5-9)</td>
<td>33A</td>
</tr>
<tr>
<td>English to Speakers of Other Languages (ESOL; grades K-12)</td>
<td>34A</td>
</tr>
<tr>
<td>Reading (grades K-12)</td>
<td>35A</td>
</tr>
<tr>
<td>Exceptional Student Education (grades K-12)</td>
<td>36A</td>
</tr>
<tr>
<td>Speech-Language Impaired (grades K-12)</td>
<td>37A</td>
</tr>
<tr>
<td>Other (please specify):</td>
<td>39A</td>
</tr>
</tbody>
</table>

**Endorsements held (check all that apply):**

<table>
<thead>
<tr>
<th>Endorsement</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>English to Speakers of Other Languages (ESOL)</td>
<td>40A</td>
</tr>
</tbody>
</table>
32  o Gifted  41A
33  o Prekindergarten Disabilities  42A
34  o Reading  43A
35  o Severe or Profound Disabilities
36  o Other (please specify): 

Did you attend the Individualizing Student Instruction professional development sessions held last spring?

37  School visit/in-service
    Y  N
38  Half-day workshop
    Y  N

When was the last class on reading and/or reading/language arts instruction you attended?

Which professional development training approaches have you attended in the past year?
Check all that apply and indicate how useful you found the approach.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Not at all Useful</th>
<th>Somewhat Useful</th>
<th>Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshops, conferences</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular collaboration with other teachers (Communities of Learning)</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual or collaborative research (action research)</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mentoring and coaching</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational visits within school</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational visits to other schools</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University courses taken for credit</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher network (including online)</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University courses other than those taken for recertification or advanced certification</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presented at workshops, training or conference</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer observation</td>
<td>1  2  3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Are you currently a member of any of the following professional organizations (check all that apply)?

- American Educational Research Association
- American Psychological Association
- Florida Reading Association
- International Dyslexia Association
- International Reading Association
- National Association for the Education of Young Children
- National Council of Teachers of English
- National Education Association
- National Federation of Teachers
- National Reading Conference
- Society for Scientific Studies in Reading
- The Council for Exceptional Children
- Other (please specify):
APPENDIX C

TEACHER FIDELITY RUBRIC
<table>
<thead>
<tr>
<th>Classroom Implementation of Individualized Instruction</th>
<th>Classroom Orient, Organization and Planning</th>
<th>Vocabulary</th>
<th>Warmth and Responsiveness/Control/Discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Rating:</td>
<td>Teacher Rating:</td>
<td>Teacher Rating:</td>
<td>Teacher Rating:</td>
</tr>
<tr>
<td>Fidelity Rating 1</td>
<td>Fidelity Rating 1</td>
<td>Fidelity Rating 1</td>
<td>Fidelity Rating 1</td>
</tr>
<tr>
<td>The teacher is not differentiating instruction.</td>
<td>The classroom is not organized. Transitions are long and instructional delivery is unclear and confusing. The general feeling is of chaos.</td>
<td>Teacher is not providing any vocabulary instruction.</td>
<td>Teacher appears as the authority figure in the class, but always punitive. Teacher does not select or incorporate students’ responses, ideas, examples, and experiences into the lesson.</td>
</tr>
<tr>
<td>Fidelity Rating 1 Indicators</td>
<td>Fidelity Rating 1 Indicators</td>
<td>Fidelity Rating 1 Indicators</td>
<td>Fidelity Rating 1 Indicators</td>
</tr>
<tr>
<td>All of the instruction is whole class without attention to individual needs of students. Instructional delivery is not appropriately paced for children of varying skill levels.</td>
<td>There is no evidence of classroom organization. Teacher frequently does not have materials ready or enough materials for all children. Classroom is frequently chaotic and very little time is spent on meaningful instruction. There is no observable system in place to facilitate students’ transition from one station or location to another.</td>
<td>Teacher does not introduce any new words. Teacher does not provide explicit or systematic instruction in vocabulary. Teacher does not provide opportunities for students to engage in oral language. Students are not encouraged to read a variety of texts inside the classroom and outside of school. Students are not provided with multiple sources of media (e.g., books on tape, graphic organizers) to facilitate understanding. Students are not provided opportunities to practice key vocabulary. Teacher does not monitor students’ vocabulary and comprehension.</td>
<td>There is no evidence that the teacher redirects in respectful ways nor is there evidence that the teacher emphasizes student change in behavior through praise. There is no evidence of the teacher communicating what students did correctly or how they can improve. There is no evidence of students treating each other with respect. Whenever discipline is imposed, it is ineffective.</td>
</tr>
<tr>
<td>Classroom Implementation of Individualized Instruction</td>
<td>Classroom Orient, Organization and Planning</td>
<td>Vocabulary</td>
<td>Warmth and Responsiveness/Control/Discipline</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Fidelity Rating 2</td>
<td>Fidelity Rating 2</td>
<td>Fidelity Rating 2</td>
<td>Fidelity Rating 2</td>
</tr>
<tr>
<td>The teacher is using primarily whole class instruction. When small groups are used, they are not always focused on literacy. Instructional delivery is inconsistently paced for children with varying skill levels</td>
<td>The classroom has inconsistent organization. Transitions are of long to reasonable duration and are not efficient. Limited instructional clarity (e.g., teachers’ instructions to children regarding how to complete activities, for example, is not always easy for students to understand).</td>
<td>Teacher provides some vocabulary instruction but it is largely defining words and, sometimes, using words in sentences. There are no opportunities for using the vocabulary in other contexts. Words are frequently Tier 1 words with few Tier 2 or 3 words selected.</td>
<td>Teacher appears as the authority figure in the class, but often punitive. Teacher rarely selects and incorporates students’ responses, ideas, examples, and experiences into the lesson. Whenever discipline is imposed, it is inconsistent and only occasionally effective.</td>
</tr>
<tr>
<td>Fidelity Rating 2 Indicators Instruction and activities are the same for all small groups.</td>
<td>Fidelity Rating 2 Indicators There may be an observable, but not efficient or working system (e.g., center chart, daily schedule) in place for organizing students into groups, inconsistent use of a lesson plan. Instructional clarity is inconsistent and children don’t always understand expectations.</td>
<td>Fidelity Rating 2 Indicators Teacher introduces too many (more than 5) new Tier 1 words per story/text and words are not any Tier 2 words. Provides explicit or systematic instruction (not both) in vocabulary. Does not extend meanings. Does not provide antonyms/synonyms or opportunities for students to engage in oral language. Students are not encouraged to read a variety of texts inside the classroom and outside of school. Students are rarely provided with multiple sources of media (e.g., books on tape, graphic organizers) to facilitate understanding. Students are rarely provided opportunities to practice key vocabulary. Teacher monitors students’ vocabulary and comprehension, but does not provide feedback.</td>
<td>Fidelity Rating 2 Indicators Teacher frequently redirects in disrespectful ways. Teacher talk is neither encouraging nor respectful. Rarely connects students’ personal experiences to lesson content. Use of directive rather than open-ended behavior management (e.g., sit down rather than everybody please go to their reading corner).</td>
</tr>
<tr>
<td><strong>Classroom Implementation of Individualized Instruction</strong></td>
<td><strong>Classroom Orient, Organization and Planning</strong></td>
<td><strong>Vocabulary</strong></td>
<td><strong>Warmth and Responsiveness/Control/Discipline</strong></td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Fidelity Rating 3</td>
<td>Fidelity Rating 3</td>
<td>Fidelity Rating 3</td>
<td>Fidelity Rating 3</td>
</tr>
<tr>
<td>Clear evidence of differentiation. The teacher is using small groups; however, the children in the small groups are generally receiving highly similar amounts and types of instruction.</td>
<td>The classroom is reasonably organized, instructional clarity is evident and transitions are fairly efficient.</td>
<td>Teacher provides fairly adequate vocabulary instruction, which may extend beyond simple definitions occasionally and provides some opportunity for using words in other contexts. There is some attempt to be intentional about selecting Tier 2 words but only about 1/3 of words contribute meaningfully to students’ understanding.</td>
<td>Teacher appears as the authority figure in the class, and is occasionally punitive. Teacher is minimally effective at selecting and incorporating students’ responses, ideas, examples, and experiences into the lesson. There is a behavior management system in place but it is used inconsistently.</td>
</tr>
<tr>
<td>Fidelity Rating 3 Indicators</td>
<td>Fidelity Rating 3 Indicators</td>
<td>Fidelity Rating 3 Indicators</td>
<td>Fidelity Rating 3 Indicators</td>
</tr>
<tr>
<td>At least one of the following is evident: Teacher managed small groups, seatwork, or child managed literacy activities. The teacher may work with a group of students while other students are participating in child managed activities. However, generally, child-managed activities are not differentiated.</td>
<td>Transitions are of reasonable length but not consistently efficient (not all children). There is an observable, but not always efficient or working system (e.g., center chart, daily schedule) in place for organizing students into groups. The teacher may use a daily lesson plan (e.g., group activity planner print-out or similar written plan).</td>
<td>Teacher introduces too many (more than 5) new Tier 1 words per story/text and not enough Tier 2 words. Provides explicit or systematic vocabulary instruction (not both). Occasionally provides antonyms/synonyms or opportunities for students to engage in oral language. Occasionally extends meanings. Students are provided occasional opportunities to practice key vocabulary. Students are encouraged to read a variety of texts inside the classroom and outside of school but are only occasionally provided with multiple sources of media (e.g., books on tape, graphic organizers) to facilitate understanding. Teacher monitors students’ vocabulary and comprehension, but rarely provides feedback.</td>
<td>Teacher inconsistently redirects in respectful ways. And inconsistently emphasizes student change in behavior through praise. Teacher talk is inconsistently encouraging and respectful and inconsistently connects students’ personal experiences to lesson content. Inconsistently communicates clearly what students did correctly or how they can improve. Students inconsistently treat each other with respect.</td>
</tr>
<tr>
<td>Classroom Implementation of Individualized Instruction</td>
<td>Classroom Orient, Organization and Planning</td>
<td>Vocabulary</td>
<td>Warmth and Responsiveness/Control/Discipline</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Fidelity Rating 4</td>
<td>Fidelity Rating 4</td>
<td>Fidelity Rating 4</td>
<td>Fidelity Rating 4</td>
</tr>
<tr>
<td>Clear evidence of differentiation. The teacher is using small groups and there is evidence that instruction is individualized. Most of the language arts block is spent in meaningful literacy activities.</td>
<td>The classroom is fairly well organized and there is adequate but not excellent instructional clarity. Instruction is usually planned in advance.</td>
<td>Teacher provides adequate vocabulary instruction, which may extend beyond simple definitions and provides some opportunity for using words in other contexts. There are attempts to be intentional about selecting Tier 2 words but only about 1/2 of words contribute meaningfully to students’ understanding and not always relevant to the text.</td>
<td>Teacher appears as the authority figure in the class, and is seldom punitive. Partially effective at selecting and incorporating students’ responses, ideas, examples, and experiences into the lesson. There is a behavior management system in place.</td>
</tr>
<tr>
<td>Fidelity Rating 4 Indicators</td>
<td>Fidelity Rating 4 Indicators</td>
<td>Fidelity Rating 4 Indicators</td>
<td>Fidelity Rating 4 Indicators</td>
</tr>
<tr>
<td>The teacher may work with a group of students while other students are participating in child managed activities. At least two of the following are evident: Teacher managed small groups, seatwork, or child managed literacy activities.</td>
<td>There is an observable, efficient, and working system (e.g., center chart, daily schedule) in place for organizing students into groups. There is an observable system in place supporting students’ transition from one station or location to another, and there is evidence that the system is working at least part of the time, the teacher consistently follows a daily lesson plan (e.g., group activity planner print-out or similar plan).</td>
<td>Teacher introduces no more than 3-5 new Tier 2 words per story/text. Provides explicit and systematic instruction in vocabulary but does not consistently extend meanings or provide antonyms/synonyms. Provides occasional opportunities for students to engage in oral language. Students are encouraged to read a variety of texts inside the classroom and outside of school but are only occasionally provided with multiple sources of media (e.g., books on tape, graphic organizers) to facilitate understanding. Students are provided occasional opportunities to practice key vocabulary. Teacher monitors students’ vocabulary and comprehension, but only occasionally provides feedback.</td>
<td>Teacher fairly frequently redirects in respectful ways and emphasizes student change in behavior through praise. Teacher talk is fairly encouraging and respectful. Fairly frequently connects students’ personal experiences to lesson content. Usually communicates clearly what students did correctly or how they can improve. Students fairly consistently treat each other with respect. There is a behavior management system in place that is used fairly consistently and effectively.</td>
</tr>
<tr>
<td>Classroom Implementation of Individualized Instruction</td>
<td>Classroom Orient, Organization and Planning</td>
<td>Vocabulary</td>
<td>Warmth and Responsiveness/Control/Discipline</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Fidelity Rating 5</td>
<td>Fidelity Rating 5</td>
<td>Fidelity Rating 5</td>
<td>Fidelity Rating 5</td>
</tr>
<tr>
<td>The teacher uses small groups, and there is good evidence that the instruction is individualized. The number and composition of groups is based on effective group size and the range of literacy skills of the students. Most of the language arts block is spent in meaningful literacy activities.</td>
<td>The classroom is well organized and there is good instructional clarity. Transitions are efficient and instructional delivery is clear.</td>
<td>Teacher provides good vocabulary instruction, which typically extends beyond simple definitions and provides students the opportunity for using words in other contexts, including written and media contexts. Selecting Tier 2 words is intentional and words contribute meaningfully to students' understanding relevant text.</td>
<td>Teacher appears as the authority figure in the class, but not punitive. Usually effective at selecting and incorporating students’ responses, ideas, examples, and experiences into the lesson. Usually effective at securing and maintaining student attention as needed. Proactively addresses behavior or effectively redirects behavior.</td>
</tr>
<tr>
<td>Fidelity Rating 5 Indicators</td>
<td>Fidelity Rating 5 Indicators</td>
<td>Fidelity Rating 5 Indicators</td>
<td>Fidelity Rating 5 Indicators</td>
</tr>
<tr>
<td>The teacher may work with a group of students while other students are participating in child managed activities. Teacher managed small groups, seatwork, and child managed literacy activities are all evident and instructional activities are in line with group mean skill levels. Instructional delivery is well paced for most of the children most of the time.</td>
<td>There is an observable, efficient, and working system (e.g., center chart, daily schedule) in place for organizing students into groups and the teacher consistently follows a daily lesson plan (e.g., group activity planner print-out or similar plan). There is an observable and efficient system in place for students to transition from one station or location to another, and there is evidence that the system is working most of the time.</td>
<td>Teacher introduces no more than 3-5 new Tier 2 words per story/text. Frequently provides explicit and systematic vocabulary instruction including extended meanings and antonyms/synonyms. Frequently provides multiple opportunities for students to engage in oral language. Students are encouraged to read and listen to a variety of texts, inside the classroom and outside of school, and are frequently provided with multiple sources of media (e.g., books on tape, graphic organizers) to facilitate understanding. Students are provided multiple opportunities to practice key vocabulary. Teacher monitors students’ vocabulary and comprehension, and provides feedback to students.</td>
<td>Teacher frequently redirects in respectful ways. Frequently emphasizes student change in behavior through praise. Teacher talk is frequently encouraging and respectful. Frequently communicates clearly what students did correctly or how they can improve. Frequently connects students’ personal experiences to lesson content. Frequently calls on a range of students. Students frequently treat each other with respect. Behavior management system used consistently and effectively.</td>
</tr>
<tr>
<td>Classroom Implementation of Individualized Instruction</td>
<td>Classroom Orient, Organization and Planning</td>
<td>Vocabulary</td>
<td>Warmth and Responsiveness/Control/Discipline</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td><strong>Fidelity Rating 6</strong></td>
<td><strong>Fidelity Rating 6</strong></td>
<td><strong>Fidelity Rating 6</strong></td>
<td><strong>Fidelity Rating 6</strong></td>
</tr>
<tr>
<td>Teachers who fully implement ISI use multiple and flexible student grouping configurations and regroup students based on formal or informal assessment data. Entire language arts block is spent in meaningful literacy activities with content of literacy instruction being differentiated.</td>
<td>The classroom is well organized and instruction is well organized. Classroom routine is evident. Transitions are efficient.</td>
<td>Teacher provides exemplary vocabulary instruction, which always extends beyond simple definitions and provides students the opportunity for using words in other contexts, including written and media and at other times of the day or week. Selecting Tier 2 words is intentional and words contribute meaningfully to students' understanding of relevant text. Tier 3 words used are appropriate.</td>
<td>Teacher is the authority figure in the class but is never punitive. Classroom consistently offers a positive learning environment with clear expectations for students' behavior as a member of the learning community. Effectively selects and incorporates students' responses, ideas, examples, and experiences into the lesson.</td>
</tr>
<tr>
<td><strong>Fidelity Rating 6 Indicators</strong></td>
<td><strong>Fidelity Rating 6 Indicators</strong></td>
<td><strong>Fidelity Rating 6 Indicators</strong></td>
<td><strong>Fidelity Rating 6 Indicators</strong></td>
</tr>
<tr>
<td>Teacher managed small groups, seatwork, and child managed literacy activities are all evident and differentiated. Students work independently at stations or are engaged in activities designed for their skill level and that reinforce developing skills while the teacher or other adult works with small homogeneous groups of students.</td>
<td>There is an observable, efficient, and working system (e.g., center chart, daily schedule) in place for organizing students into groups and teacher consistently follows a daily lesson plan (e.g., group activity planner print-out or similar written plan). Activities run smoothly with few disruptions. Materials are readily available to students. Disruptions are handled quickly and effectively. Students know what to do and what is expected of them. There is an observable, efficient, and working system in place for students to transition from one station or location to another and is evidence that the system is working well.</td>
<td>Teacher introduces no more than 3-5 new Tier 2 words per story/text. Consistently provides explicit and systematic instruction in vocabulary, including extended meanings and antonyms/synonyms. Consistently provides multiple opportunities for students to engage in oral language. Students are encouraged to read and listen to a variety of texts, inside the classroom and outside of school, and are consistently provided with multiple sources of media (e.g., books on tape, graphic organizers) to facilitate understanding. Students are provided multiple opportunities to practice key vocabulary. Teacher monitors students’ vocabulary and comprehension, and provides feedback to students.</td>
<td>Teacher consistently redirects behavior in respectful ways, using friendly or unobtrusive, respectful gestures to redirect. Teacher talk is consistently encouraging and respectful, emphasizing student change in behavior through specific praise. Behavior management system is used consistently and effectively. Consistently communicates clearly what students did correctly or how they can improve. Consistently connects students’ personal experiences to lesson content, calling on a range of students and eliciting responses from all students, including those having difficulty with task at hand. Frequently encourages peer support in the learning process. Students consistently treat each other with respect.</td>
</tr>
</tbody>
</table>
Dear Teachers,

**How can we be more effective teaching students how to read?**

For the past several years, this question has been at the center of our research. We are writing to invite you to participate in the Individualizing Student Instruction (ISI) project, which is designed to find out whether providing optimal child-specific amounts and types of literacy instruction to all students, not just struggling readers, based on students’ vocabulary and reading skills, will ensure that all students in the classroom are reading at or above grade level by the end of the school year according to Florida’s state standards. There are two interventions, ISI-Reading and the Teacher Study Group. You will be assigned, at random, to one of these intervention groups. At the end of the study, you will receive training in the intervention to which you were not assigned.

Teachers in the reading intervention (ISI-reading) will receive training on how to provide individualized reading instruction and how to use Assessment-to-Instruction (A2I) software designed to help teachers plan and provide amounts and types of instruction for each student based on his or her entering reading and vocabulary skills. Teachers in the Teacher Study Group will learn how to teach evidence-based activities designed to improve students’ vocabulary skills.

Participating teachers will receive technology designed to help them implement ISI-Reading or the Vocabulary interventions plus $500 compensation per year – half in the middle and half at the end of the school year. You are agreeing to participate in a one year study and, we hope, you will continue to provide the interventions at the end of the study year but this is not required. We will share all the results of the research with you as they are available.

You can expect the following during the year of the study:

- **Videotaping your classroom at least three times per year:** during language arts instruction in the fall and spring, and all-day in the winter. You may be asked to video tape selected vocabulary or reading activities. The purpose of the all-day videotaping is to capture all of the other opportunities throughout the school day where language arts, reading and vocabulary instruction occurs. Portions of this tape may be used for training and educational purposes. This is a very important part of the study.

- **Assessing your students’ reading, vocabulary, and other key skills at least three times during the school year.** This information will be used to design individualized reading instruction, follow students’ progress, and to improve the A2I software and vocabulary interventions.

- **Asking you to complete several questionnaires.**

- **During the study year, you will receive training, including classroom support, about once every other week if you are in the ISI-Reading intervention. The Teacher Study Group teachers will meet monthly. Interactions may be more frequent in the fall. You will also have an opportunity to interact with other teachers receiving the training in small and large groups throughout the school year.**

- **During the study, we ask you to follow the protocols carefully. This includes for the ISI-reading intervention, using the A2I software at least every other week for about 20 minutes to plan your instruction. Our research suggests that this is the minimum amount needed to help your students achieve educationally important gains in reading. For teachers in the Teacher Study group, this means implementing the vocabulary activities you have designed and sharing the activities with the other teachers in the Teacher Study Group.**
• All information will be kept in secure locations, and all digital files will be password protected so that only project researchers have access to it. All information will be kept for at least seven years, as required by the grant, in secure locations, and will be destroyed when it is no longer needed.

Your participation is voluntary and you are free to withdraw your consent at any time without penalty or prejudice. You will be able to skip any questions on the questionnaires that you do not want to answer. Any new information that might develop during this project will be provided to you if it might affect your willingness to participate in the project. You will not be identified in any reports on this study. Records will be kept confidential to the extent provided by federal, state and local law. The sponsor of the project, the US Department of Education, or other official agencies responsible for monitoring this study may inspect these records.

If you have any questions now or in the future, please contact me at the Florida Center for Reading Research, (850) 921-0703 or by email, cconnor@ferr.org. You may also contact the Project Director, Dr. Elizabeth Crowe by email ecrowe@gmail.com or by phone 850-645-2980. Should you have questions regarding your rights as a participant in research, please contact the Behavioral Sciences Institutional Review Board, (850) 644-8633.

We hope you will become involved. This study will help us develop better ways to teach reading and math, help children learn, and find out what we can do to make sure all children succeed in school.

If you are willing to participate in this study, please sign below. A copy of this letter is provided for your records. Thank you so much.

Sincerely,

Carol McDonald Connor, Ph.D.
Assistant Professor, FSU College of Education and Florida Center for Reading Research

I am willing to participate in the study entitled: Child-Instruction Interactions in Early Reading: Examining Causal Effects of Individualized Instruction. I understand that participation is voluntary and that I can withdraw from the study at any time. I understand that activities in my classroom will be videotaped during the study, that I will participate in professional development activities, and will do my best to use what I have learned when instructing my students. I also understand that I will receive $200 compensation, supporting technology the A2I-Reading software, when available, and the vocabulary lessons designed by the Teacher Study Group.

Signature: __________________________________________

Printed Name: _______________________________________

Date: _______________________________________________
Dear Parents,

How can we be more effective teaching our children how to read and do mathematics?

For the past several years, this question has been at the center of our research. I am writing to invite you and your child to participate in a three year study of how to teach students reading and mathematics better. Our research shows that the effect of instruction depends on children's reading, language and other skills. If we tailor instruction to meet each child's needs, children learn better. Your child's teacher this year is participating in a research project designed to help teachers provide more effective reading and mathematics instruction. We have attached an article on the reading intervention. The math intervention will work much the same way but we will be giving teachers information on paper rather than through the software. We are asking all of the students in your child's class to participate. Your child's teacher this year will receive training in how to improve students' reading, math, and language skills and how to individualize this instruction using assessment results. Then your child's second grade and third grade teachers will receive the training. From first through third grade, we will be following your child's progress learning to read and do mathematics by assessing your child's mathematics and reading skills and video-taping during the school day, including reading and mathematics instruction. If you give permission, below is what we will be doing and when.

- In order to see how well our training is working, we will be videotaping your child's classroom at least 3 times per year. These tapes will be used for education and research purposes including as feedback to help teachers individualize reading and mathematics instruction and to understand the way that they teach reading and math. This videotaping is a very important part of our study. If you do not give permission, we will blur your child's identity on these tapes.
- We will assess your child in the fall, winter and spring for about 30 minutes per visit, each year. During these visits, we will assess the participating students in your child's class using a variety of language, reading, mathematics, and thinking assessments. The results of these assessments will be shared with the teachers as part of the study. Reading and mathematics results will be provided to you as well. If you do not give permission, your child will continue to receive instruction but will not receive an updated individual plan based on the assessment results.
- You will be asked to complete questionnaires seeking information about your child and family. You do not need to answer any questions that you do not want to answer.
- All of this information will be kept in secure locations and all digital files will be password protected so that only project researchers can access it. All information, including videos, will be kept for at least seven years, as required by the grant, in a secure location and only project staff will have access to it. The information will be destroyed when it is no longer needed.
- We will use assessment information (including DIBELS, PPVT, & SAT10 mathematics and reading) about your child that is collected by school district teachers and staff. Some of this information will be accessed through the Progress Monitoring Reporting Network (PMRN) and through district records using your child's identification numbers and we will continue to monitor your child's progress on these measures.

As compensation for your time and effort, you will receive a $20 gift card at the end of the school year. This study has the approval of the Bay County School District and will last for the entire school year: August through May. And then the following years until your child completes third grade. Your participation is voluntary and you are free to withdraw your consent at any time without penalty or prejudice. Any new information that might develop during this project will be provided to you if it might affect your willingness to participate in the project. We will also send you a letter at the beginning of each school year that will provide an overview of the project and results thus far. It will also remind you of the voluntary nature of this study and your and your child's participation.

If you have any questions now or in the future, please contact me at Florida State University at 850-921-0703 or by email, cconnor@fcrr.org, or call the project director, Elizabeth Crowe at 850-830-3799. Should you have questions regarding your rights as a participant in research, please contact the FSU Institutional Review Board, (850) 644-8633.

We hope you will become involved. This study will help us develop better ways to teach reading and math, how to help children learn, and discover what we can do to make sure all children succeed. Please complete the enclosed form in the spaces provided and return it using the envelope provided. The extra copy is for you. Thank you so much.

Sincerely,

Carol McDonald Connor, Ph.D. Associate Professor, FSU College of Education and Florida Center for Reading Research

August 13, 2008
FSU/Florida Center for Reading Research

Individualizing Student Instruction Project

Please complete the following information, sign in the appropriate space below, and return in the enclosed postage paid reply envelope. Please keep one copy for yourself. Should you have questions regarding your rights as a participant in research, please contact the FSU Institutional Review Board (850) 644-8633 or contact the Florida Center for Reading Research, (850) 644-9352.

Child’s name: ___________________________ Parent(s) Name: ___________________________

Relation to Child: ___________________________ Child’s date of birth: ___________________________

Home phone number: ___________________________ Address: ___________________________

Street

City ___________________________ Zip ___________________________

Child’s School: ___________________________ Child’s Teacher: ___________________________

Please read the following statement and sign below if you are willing for your child to participate in the study.

If you check, “I give my permission for my child to participate in this study” this means that your child’s reading or math instruction will be tailored to his or her level of reading or math skill based on assessment results and your child’s teacher will receive training in how to do this, as well as materials and information to support this individualized instruction. I understand that our participation is voluntary and that we can withdraw from the study at any time and that I may ask any survey question that I do not wish to answer. I understand that as part of this study, part of the school classroom visits will be videotaped and is a requirement of the study. I understand that neither my child nor I will be identified in any published reports on this study. Records will be kept confidential to the extent provided by federal, state, and local law. I will receive a $20 gift card at the end of the school year as compensation for our family’s participation. The sponsors of the project, the US Department of Education and NICHD, or other official agencies responsible for monitoring this study may inspect these records.

If I do not give permission for my child to be in the study by checking the “no” box below or not returning the form, the teacher will provide the reading and math instruction but will not receive an updated instruction plan for my child based on assessments and my child will not be assessed.

In order to see how well our training is working, we will be videotaping your child’s classroom at least 3 times per year. If I do not give consent for my child to be in the ISI study or if I wish to withdraw from the study then my child’s identity will be digitally blurred on the videotapes only if I did not give media permission to Bay County. Bay County School district sent home a media permission form for your child to appear in photos and videos at the beginning of the year. In any case, my child’s identity will be protected and the tapes will be kept in a secure location seen only by research staff. If I have any questions about the video taping and want to make other arrangements, I may call Carol Connor at 850-921-0703.

I give permission for my child to participate.

☐ Signed: ___________________________ Date: ___________________________

☐ I give permission for my child to participate, and I have questions. Please call me.

Signed: ___________________________ Date: ___________________________

☐ No, I am not interested in participating at this time.

Signed: ___________________________ Date: ___________________________
RE- APPROVAL MEMORANDUM Date: 5/13/2011 To: Carol Connor Address: PDB 306- 4304

Dept.: LEARNING SYSTEMS INSTITUTE From: Thomas L. Jacobson, Chair

Re: Re- approval of Use of Human subjects in Research Examining Child- Instruction Interactions in Early Reading: Examining causal effects of individualized instruction

Your request to continue the research project listed above involving human subjects has been approved by the Human Subjects Committee. If your project has not been completed by 5/9/2012, you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the committee.

If you submitted a proposed consent form with your renewal request, the approved stamped consent form is attached to this re-approval notice. Only the stamped version of the consent form may be used in recruiting of research subjects. You are reminded that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report in writing, any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor are reminded of their responsibility for being informed concerning research projects involving human subjects in their department. They are advised to review the protocols as often as necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

Cc: HSC No. 2011.6332Human Subjects <humansubjects@magnet.fsu.edu> Friday, May 13, 2011 2:38 PM cconnor@fsu.edu Use of Human Subjects in Research - Approval Memorandum
BIBLIOGRAPHY


Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33: 159-74.


BIOGRAPHICAL SKETCH

EDUCATION

Ph.D. Florida State University, Reading Education and Language Arts (2012)
M.S. Nova Southeastern University, Elementary Education (2007)
B.A. Florida State University, English Literature (2004)

RESEARCH AND TEACHING

Florida Center for Reading Research Tallahassee, FL. (2009-2010)
Part-time Assistant in Research (Graduate Assistant)
- Co-authored manuscript concerning third grade reading comprehension
- Analyzed student observational data to identify instances of individualized student instruction
- Coded student video data utilizing NoldusXT Observer Coding Software program
- Developed mathematics-based coding system for iSi Math
- Served as a graduate teaching assistant to Dr. Carol Connor for Child Psychology course

Florida Center for Reading Research Tallahassee, FL. (2008-2009)
Full-time Faculty, Assistant-in Research (Master Teacher)
- Co-authored published manuscript in the peer-reviewed journal, Reading Teacher
- Implemented the Individualizing Student Instruction (iSi) Science Intervention
- Taught 2 complete Science units in 6 second grade classrooms
- Developed science curriculum
- Conducted monthly Community of Practice Meetings with the teachers in the iSi Science study
- Developed and administered assessments
- Managed student data

Florida Center for Reading Research Tallahassee, FL. (2007-2008)
Full-time Faculty, Assistant in Research (Professional Development Partner)
- Developed differentiated instruction and classroom management strategies for teachers
- Trained teachers on how to effectively implement individualized instruction for all students in their classroom
Facilitated the use of study-related software (A2i) in an effort to promote the use of individualized student instruction
Recruited third grade teachers for participation in a longitudinal study
Recorded field notes, observed classroom practices, managed data, assessed students, assisted with videography, and created FCRR-center activities
Developed teacher training workshops

Gulliver Academy-Gulliver Schools Miami, FL. (2006-2007)
- Taught First Grade to a classroom on 18 students utilizing Open Court reading and language arts curriculum and D’Nelian handwriting
- Utilized an online audio portfolio to record and track students’ oral reading fluency
- Created developmentally appropriate practices for students in daily classroom activities
- Implemented student center activities as a way to support reading growth and interest

Gulliver Academy-Gulliver Schools Miami, FL. (2005-2006)
- Taught Kindergarten to a classroom of 21 students utilizing Open Court reading and language arts curriculum and D’Nelian handwriting
- Integrated center activities and rotations into daily classroom routine
- Fostered a developmentally appropriate classroom learning environment

- Taught weekly two-hour classes to a group of children ranging in age from 4 to 8 years of age
- Stimulated students’ learning by encouraging their creativity and critical thinking skills
- Taught students’ effective communication skills, leadership abilities, and encouraged cooperative learning skills
- Developed students’ verbal and non-verbal intelligence
- Fostered students’ goal setting, attentiveness and concentration skills

Smart Starts Academy Miami, FL. (2004-2005)
- Taught preschool to a classroom of 24 students utilizing the Abeca curriculum and Zaner Bloser handwriting
- Integrated center activity time daily to promote learning through exploration
- Created an exciting learning environment for students
- Fostered knowledge of print concepts
- Taught alphabetic principles, numbers, and colors

TUTORING
2009-2010  Tutored one 3rd. grade student in writing fluency and reading comprehension for one hour weekly
2008-2009  Tutored two 1st. grade ESOL students weekly for an hour in reading, writing, spelling, and vocabulary skills
2008-2009  Tutored one 2nd. grade student in word recognition and reading comprehension skills
2007-2008  Tutored one 6th grade student weekly for 2 hours in reading and writing skills
2006-2007  Tutored two students, one 1st. grader and one 3rd. grader for an hour each in reading

QUALIFICATIONS

- Professionally Certified Florida Educator: Pre-Kindergarten-3rd, English Grades 6-12, Elementary Education K-6, and Reading and Language Arts K-12
- 5 years Teaching Experience
- 3 years Education Research Experience
- Completed 45 hours Child Care Training at Miami Dade College, Awarded CDA
- Certified CPR and First Aid, AED Administrator