

6-2-2008

# Teacher Efficacy and Student Achievement in Ninth and Tenth Grade Reading: A Multilevel Analysis

Anete Vasquez  
*University of South Florida*

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

---

## Scholar Commons Citation

Vasquez, Anete, "Teacher Efficacy and Student Achievement in Ninth and Tenth Grade Reading: A Multilevel Analysis" (2008).  
*Graduate Theses and Dissertations*.  
<https://scholarcommons.usf.edu/etd/544>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Teacher Efficacy and Student Achievement in Ninth and Tenth Grade Reading:  
A Multilevel Analysis

by

Anete Vasquez

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Secondary Education  
College of Education  
University of South Florida

Major Professor: Jane Applegate, Ph.D.  
Roger Brindley, Ph.D.  
Patricia L. Daniel, Ph.D.  
Melinda Hess, Ph.D.  
Kofi Marfo, Ph.D.

Date of Approval:  
June 2, 2008

Keywords: adolescent literacy, ohio state teacher efficacy scale, hierarchical linear modeling, confirmatory factor analysis, standardized testing

© Copyright 2008, Anete Vasquez

## DEDICATION

This work is dedicated to my parents, Asja Vásquez and Guillermo Alejandro Vásquez, who have instilled in me the importance of education, the inspiration to set high goals and the confidence to achieve those goals. It is also dedicated to Bryan J. Williams who offered serenity and support throughout the writing of this dissertation.

## ACKNOWLEDGEMENTS

It is with deep appreciation that I would like to acknowledge those individuals who have given their time, expertise and energy to help me complete this work. First, I thank Dr. Jane Applegate, my major professor, whose willingness to read multiple drafts, provide insightful feedback and encourage me to stay the course made writing this dissertation more about the journey than the destination. I have arrived, and I am wiser because of lessons learned along the way. I would also like to extend gratitude to my committee members Dr. Roger Brindley, Dr. Patricia L. Daniel, Dr. Melinda Hess and Dr. Kofi Marfo. From each of you I have learned different aspects of what it means to be a scholar. Dr. Bethany Bell-Ellison not only provided invaluable assistance with the statistical analysis, she also offered encouragement when it was needed most. I am also indebted to the directors and staff of the Offices and Research and Accountability, the English / Language Arts supervisors and teachers in the two participating school districts.

## TABLE OF CONTENTS

LIST OF TABLES .....	iv
LIST OF FIGURES .....	vii
ABSTRACT .....	viii
CHAPTER 1: INTRODUCTION .....	1
1.0 Context of the Problem .....	5
1.1.1 Professional Organizations in Support of Adolescent Literacy .....	6
1.1.2 Federal Support for Adolescent Literacy .....	12
1.1.3 The Push for Highly Qualified Teachers .....	15
1.1.4 Teacher Effectiveness and Teacher Efficacy .....	19
1.2 Statement of the Problem .....	21
1.3 Purpose of the Study .....	24
1.4 Research Questions .....	24
1.5 Methodology .....	24
1.6 Conceptual Framework .....	26
1.7 Significance of the Study .....	27
1.8 Basic Assumptions .....	29
1.9 Limitations .....	31
1.10 Definition of Terms .....	34
1.11 Organization of the Dissertation .....	36
1.12 Summary .....	37
CHAPTER 2: REVIEW OF THE LITERATURE .....	40
2.1 Adolescent Literacy .....	41
2.1.1 The Development of Adolescent Literacy .....	40
2.1.2 Current Trends in the Study of Adolescent Literacy .....	44
2.1.3 Summary of Adolescent Literacy .....	47
2.2 Teacher Beliefs .....	48
2.2.1 Teacher Efficacy in Albert Bandura’s Social Cognitive Theory .....	55
2.2.2 Sources of Teacher Efficacy .....	57
2.2.3 Methods of Measuring the Construct of Efficacy .....	60
2.2.4 The Teacher Efficacy Scale and Problems of Interpretation .....	64
2.2.5 Summary of Teacher Beliefs .....	66

2.3	The Ohio State Teacher Efficacy Scale .....	68
2.3.1	Summary of The Ohio State Teacher Efficacy Scale .....	75
2.4	Teacher Efficacy and its Relationship to Various Aspects of Schooling .....	75
2.4.1	Teacher Efficacy and Student Achievement in Reading .....	76
2.4.2	Summary of Teacher Efficacy and its Relationship to Various Aspects of Schooling .....	81
2.5	The Three Subscales of The Ohio State Teacher Efficacy Scale.....	82
2.5.1	Teacher Efficacy for Student Engagement .....	83
2.5.2	Teacher Efficacy for Instructional Strategies .....	86
2.5.3	Teacher Efficacy for Classroom Management .....	92
2.5.4	Summary of the Three Subscales of The Ohio State Teacher Efficacy Scale .....	97
2.5.5	Chapter Summary .....	99
CHAPTER 3: METHODOLGY .....		101
3.1	Purpose of the Study .....	101
3.2	Research Questions.....	102
3.3	Research Design.....	102
3.3.1	Sample.....	102
3.3.2	Data Collection .....	115
3.3.3	Data Management .....	116
3.3.4	Dependent Variable .....	117
3.3.5	Independent Variables .....	120
3.3.6	Contextual Variables Taken into Consideration .....	122
3.4	Data Analysis .....	127
3.4.1	Advantages of Hierarchical Linear Modeling (HLM) .....	127
3.4.2	Characteristics of HLM.....	129
3.4.3	Methodological Points Related to the Use of HLM.....	131
3.4.4	The HLM Models .....	134
3.4.5	Interpretation Techniques .....	149
3.5	Summary .....	149
CHAPTER 4: RESULTS.....		151
4.1	Purpose of the Study .....	151
4.2	Research Questions.....	152
4.3	Descriptive Information .....	153
4.3.1	Sample.....	153
4.4	Confirmatory Factor Analysis of The Ohio State Teacher Efficacy Scale .....	159
4.5	Checking Assumptions for the HLMs .....	163
4.6	The Two-Level HLM Analysis.....	167
4.6.1	Research Question 1 .....	175
4.6.2	Research Question 2 .....	177
4.6.3	Research Question 3 .....	179
4.6.4	Research Question 4 .....	181

4.7 Summary of Findings.....	183
CHAPTER 5: DISCUSSION.....	184
5.1 Background of the Study .....	184
5.2 Purpose of the Study .....	187
5.3 Research Questions.....	188
5.4 Discussion of Findings.....	188
5.5 Limitations of the Study.....	191
5.6 Implications.....	193
5.7 Recommendations for Future Research .....	194
5.7.1 Recommendations for Researcher in Methodology.....	194
5.7.2 Recommendations for Research in Policy .....	197
5.7.3 Recommendations for Research in Teacher Preparation and Professional Development .....	202
5.8 Unanswered Questions.....	204
5.9 Conclusions.....	212
REFERENCES .....	214
APPENDICES .....	233
Appendix A: Researcher Script for Department Chairs’ Monthly Meeting .....	234
Appendix B: Directions for Administering the Survey .....	235
Appendix C: Letter to Teacher Participants.....	236
Appendix D: Teacher Personal Information Questionnaire.....	237
Appendix E: Teachers’ Sense of Efficacy Scale.....	238
Appendix F: Informed Consent Form.....	239
Appendix G: Additional Confirmatory Factor Analysis Tables.....	241
Appendix H: Assumption Checking for the HLMs .....	246
Appendix I: Summary Table of the HLMs.....	258
Appendix J: An Explanation of Developmental Scale Scores.....	270
ABOUT THE AUTHOR .....	275

## LIST OF TABLES

Table 1:	Review of Research on Teacher Efficacy and Student Achievement in Reading .....	80
Table 2:	Demographic Information for Schools of Teacher Participants .....	104
Table 3:	Correlations between Teacher Efficacy Scores and Years Teaching and Age .....	105
Table 4:	Efficacy Score Means by Degree Achieved .....	106
Table 5:	Efficacy Score Means by Type of Certification .....	107
Table 6:	Efficacy Score Means by Gender .....	107
Table 7:	Efficacy Score Means by Ethnicity.....	108
Table 8	Reading Developmental Scale Scores (86 to 3008) for each Achievement Level on the Florida Comprehensive Assessment Test for Ninth and Tenth Grade .....	119
Table 9:	Summary of the Model Structure for each Hierarchical Linear Model .....	137
Table 10:	Student Demographic Information .....	154
Table 11:	Florida Comprehensive Assessment Test Developmental Scale Scores for Ninth and Tenth Grade Students .....	155
Table 12:	Reading Developmental Scale Scores (86 to 3008) for each Achievement Level on the Florida Comprehensive Assessment Test for Ninth and Tenth Grade .....	157
Table 13:	Teacher Demographic Data for Confirmatory Factor Analysis and Hierarchical Linear Models .....	158
Table 14:	Mean Scores for Ohio State Teacher Efficacy Scale .....	160
Table 15:	Bivariate Correlation Matrix for all Criterion and Predictor Variables .....	165



Table 16:	Distribution of Level-1 and Level-2 Residuals.....	166
Table 17:	Fixed and Random Effects for the 2-Level Unconditional Model .....	169
Table 18:	Hierarchical Linear Models with Level-1 Variables Added.....	174
Table 19:	Hierarchical Linear Model of Best Fit for Instructional Strategies .....	176
Table 20:	Hierarchical Linear Model of Best Fit for Classroom Management .....	178
Table 21:	Hierarchical Linear Model of Best Fit for Student Engagement .....	180
Table 22:	Hierarchical Linear Model of Best Fit for Total Efficacy .....	182
Table G1:	Correlation Matrix of the Three-factor Model.....	241
Table G2:	Standardized Residuals of the Confirmatory Factor Analysis.....	243
Table G3:	Parameter Estimates from the Three-Factor CFA Models of Efficacy in Student Engagement, Instructional Strategies and Classroom Management.....	244
Table I1:	HLMs for Instructional Strategies with Level-1 Variables.....	258
Table I2:	HLMs for Instructional Strategies with Two-Way Main Effects and Interactions.....	259
Table I3:	HLMs for Instructional Strategies with Three-Way Main Effects and Interactions.....	260
Table I4:	HLMs for Classroom Management with Level-1 Variables.....	261
Table I5:	HLMs for Classroom Management with Two-Way Main Effects and Interactions.....	262
Table I6:	HLMs for Classroom Management with Three-Way Main Effects and Interactions.....	263
Table I7:	HLMs for Student Engagement with Level-1 Variables .....	264
Table I8:	HLMs for Student Engagement with Two-Way Main Effects and Interactions.....	265
Table I9:	HLMs Student Engagement with Three-Way Main Effects and Interactions.....	266

Table I10:	HLMs for Total Efficacy with Level-1 Variables.....	267
Table I11:	HLMs for Total Efficacy with Two-Way Main Effects and Interactions.....	268
Table I12:	HLMs Total Efficacy with Three-Way Main Effects and Interactions.....	269

## LIST OF FIGURES

Figure 1.	Definition of scientifically based research in No Child Left Behind Act of 2001 .....	198
Figure H1.	Box-and-whisker plot for Level-1 residuals (instructional strategies).....	246
Figure H2.	Box-and-whisker plot for Level-2 residuals (instructional strategies).....	247
Figure H3.	Box-and-whisker plot for Level-1 residuals (classroom management) .....	248
Figure H4.	Box-and-whisker plot for Level-2 residuals (classroom management) .....	249
Figure H5.	Box-and-whisker plot for Level-1 residuals (student engagement).....	250
Figure H6.	Box-and-whisker plot for Level-2 residuals (student engagement).....	251
Figure H7.	Box-and-whisker plot for Level-1 residuals (total efficacy).....	252
Figure H8.	Box-and-whisker plot for Level-2 residuals (total efficacy).....	253
Figure H9.	Level-2 Instructional Strategies residuals*predicted reading achievement .....	254
Figure H10.	Level-2 Classroom Management residuals*predicted reading achievement .....	255
Figure H11.	Level-2 Student Engagement residuals*predicted reading achievement. ....	256
Figure H12.	Level-2 Total Efficacy residuals*predicted reading achievement .....	257

# TEACHER EFFICACY AND STUDENT ACHIEVEMENT IN NINTH AND TENTH GRADE READING: A MULTI-LEVEL ANALYSIS

Anete Vásquez

## ABSTRACT

More than 8 million of America's middle and high school students are struggling readers. Two-thirds of all eighth graders read below grade level, and the reading scores of high school students have not improved since 1974. Low literacy levels affect learning in all subject areas and impede student opportunity for future success. The No Child Left Behind Act of 2001 and the accountability measures associated with the Act have heightened public awareness of the deficiency in adolescent literacy. School districts are choosing to respond in one of two ways. Some school districts are opting to invest in teacher-proof curricula that negate the effect of the teacher. Other districts are opting to invest in the professional development of their teachers.

The goal of this study was to support district efforts to provide strategic professional development opportunities for teachers by investigating the effects of teacher efficacy for instructional strategies, classroom management and student engagement on ninth and tenth grade students' reading achievement. Teachers with high efficacy were hypothesized to impact students' reading gains positively. Student contextual variables of prior achievement, socioeconomic status, ethnicity and grade were controlled for in the study. The participants included 2,061 students in 23 classrooms taught by 110 teachers in two school districts on the west coast of Florida. The results

indicated that there was no statistically significant relationship between teacher efficacy and student reading achievement gains. The only variables of statistical significance were race (White vs. non-White) and grade.

As more researchers use the findings and recommendations from this study to inform new investigations of the complex relationship between teacher efficacy and student achievement in reading, teacher educators, policymakers, teachers and administrators will be better informed as they continue to work towards improving the reading achievement scores of and narrowing the achievement gaps in adolescent literacy.

## CHAPTER 1

### INTRODUCTION

Student achievement in reading is receiving nation-wide attention. While many factors have been examined to try to understand more about adolescent reading development and how to improve adolescent literacy, teachers are feeling hard-pressed to raise the achievement levels of their students. Consider, now, the following scenario:

Katherine Jewel put down her Saturday newspaper in frustration. The headline read “FCAT-based bonus proposed,” and the article outlined a performance-based pay structure for teachers. Unlike many teachers at her school, Ms. Jewel did not see the state’s comprehensive examination as “the enemy.” She believed in accountability, and she actually looked forward to receiving her students’ scores during the second week of each May. Furthermore, she believed that her state’s comprehensive assessment test was a good measure of students’ abilities. Therein, however, was the problem as well; it was a good measure – one measure. Ms. Jewel believed in the definition of assessment that she had learned from her methods courses at the college of education she attended. In that definition, assessment was defined as “a set of strategies for discovering what students know or can do as a result of engaging in learning experiences. It involves a number of activities designed to determine the level of student learning” (Knowles & Brown, 2000).

She thought that the comprehensive exam was a good measure that should be coupled with many other measures before making statements about student ability.

Ms. Jewel also did not like what was being done with the exam with regard to teachers and schools. As far as teachers were concerned, Ms. Jewel felt too much pressure was placed on teachers to “teach to the test.” While the state department of education released an informative booklet about the test that clearly stated that “teachers should never cease instruction in their ordinary lessons to spend time teaching the content required by the [exam]” and that “no school should ignore its responsibilities or spend long hours in activities called [test] preparation,” that is exactly what most of the schools in her district did (Florida Department of Education, 2004, p. 20). As a matter of fact, the language arts supervisor for Ms. Jewel’s entire county distributed a three-month calendar outlining test preparation activities that each language arts teacher was supposed to accomplish daily in the weeks preceding the February test dates. At her own school, the principal required her department to create non-fiction reading activities for the other subject area teachers to complete with their students. None of this was well received by the faculty.

The teachers of the language arts department were the most disheartened. Many of them felt that the majority of the weight of the school grade, assigned to the school by the state, rested on their shoulders because the exam was comprised of three sections: mathematics, reading and writing. Soon there would be a science section as well, but as it now stood, language arts teachers prepared students for two-thirds of the test. Adding to the discontent of her department was the fact that the language arts teachers knew that the

district-mandated reading activities geared specifically to the preparation for the exam did not turn students into lifelong readers and lovers of literature. Ms. Jewel thought of her own classes and the sighs students emitted when she pulled out the exam preparation books purchased by her district. “Not the *Eureka* books,” they would moan.

Ms. Jewel also wondered at the time, energy and money that went into all of this test preparation. She remembered when she first started teaching the district council of teachers of English sponsored amazing workshops on teacher training days, and the district supervisor used to have time to visit school sites to see how teachers new to the district were doing. Neither of those events occurred since the advent of the mandated testing. She thought of her own colleagues. Some of them motivated their students to achieve at amazing levels; others struggled. This was evident to everyone in the language arts department because students’ mean scores were displayed alongside teachers’ names at department meetings. Although many teachers worked with very similar student populations, the results those teachers achieved varied dramatically. Ms. Jewel’s own students progressed at a moderate level. The teacher two portables down from her, Ms. Powers, worked nothing short of miracles with her students. When questioned about her strategies, Ms. Powers’ replies were not specific enough to satisfy Ms. Jewel. Time-constraints of the academic day and Ms. Power’s own modesty prevented Ms. Jewel from discovering how Ms. Powers got her students to attain such high yearly learning gains. Highest student achievement was a mantra of the administrators at her school, and teachers were held accountable, especially in the language arts department.



This story is typical of most schools in most states across the nation. Teachers and administrators feel a great deal of pressure with regard to improving the standardized test scores of all students. This pressure is heightened by the threat of mandated school restructuring if schools do not meet adequate yearly progress (AYP) indicators as outlined by the *No Child Left Behind (NCLB) Act of 2001*. Depending upon the percentage of AYP indicators not met, school restructuring can result in major restructuring of the school's governance, schools turning their operations over to the State for monitoring or entering into a contract with a private management company with a demonstrated record of effectiveness in operating schools and reopening the school as a public charter school. Measures can also include replacing all or most of the school staff, including the principal, who are relevant to the failure to make AYP. This last restructuring mandate requires school districts to keep data on all school personnel and their contributions toward the goal of meeting AYP (Florida Department of Education, *NCLB School Restructuring*, 2007). Thus, it is understandable why there is now a great deal of emphasis on improving students' standardized test scores. Secondary schools are particularly concerned with reading scores, as adolescent reading scores have shown little improvement over the years (National Center for Educational Statistics [NCES], 2004). While the pressure of high stakes testing is relatively new in many states, the question of how to improve adolescent literacy, however, is not a new dilemma.

In 1976, a group of RAND researchers studied reading programs in Los Angeles. They found that while teacher background seemed to have little relationship to student gains in reading, teachers' sense of efficacy, or their belief that "he or she has the

capacity to produce an effect on the learning of students,” was strongly correlated to student achievement in reading (Armour et al., 1976). In the study, two simple questions were posed to teachers. The first statement on the questionnaire, “when it comes right down to it, a teacher really can’t do much because most of a student’s performance depends on his or her home environment,” and the second, “if I try really hard, I can get through to even the most difficult or unmotivated students,” were combined to create a measure of teacher efficacy. It was discovered that teachers with a higher sense of efficacy also had students who achieved at higher levels with regard to reading achievement. The findings of the RAND researchers garnered a great deal of interest in the concept and promise of teacher efficacy. Thirty years later, the concept of efficacy is still an item of interest. As recently as 1990, Woolfolk and Hoy noted that although “researchers have found few consistent relationships between characteristics of teachers and the behaviors or learning of students, teachers’ sense of efficacy is an exception to this general rule” (p. 81). Despite the fact that teacher efficacy has been a highly studied construct in relation to student achievement in general, little more has been done to investigate the relationship of teacher efficacy to student achievement in reading.

### 1.1 Context of the Problem

Student achievement in reading, however, is a topic that is currently receiving much attention because the reading scores of America’s middle school and high school students have shown no improvement at all between the years of 1971 and 2004 (NCES, 2004). As a matter of fact, the average reading scores for 9-year-old students were at an all time high in 2004; scores for 13-year-old students have risen only three points since

1975, and scores for 17-year-old students have dropped five points since 1992 (Perie, Moran, & Lutkus, 2005).

### *1.1.1 Professional Organizations in Support of Adolescent Literacy*

The fact that many adolescents struggle with reading is an issue that many acknowledged; the fact that little was being done to rectify the situation was largely unchallenged until the late 1990s. Shortly after serving as president of the International Reading Association (IRA), Richard Vacca wrote:

I am beginning to wonder whether there is a political and public mindset that literacy learning is critical only in early childhood. The faulty and misguided assumption, 'If young children learn to read early on, they will read to learn throughout their lives,' results in more harm than good (1998, p.606).

In recent years, influential professional organizations have begun to draw attention to the needs of adolescent readers. In 1997, the IRA formed the Adolescent Literacy Commission (ALC) and undertook several initiatives to examine the state of adolescent literacy nationwide. The ALC sponsored conference sessions and institutes in major cities to examine the challenges faced by secondary teachers with regard to literacy issues and to highlight exemplary practices and programs occurring in secondary schools across the nation. The ALC also began a campaign to increase public support for adolescent literacy programs. One of the most influential papers to emerge from this campaign was the IRA's *Position Statement on Adolescent Literacy* (Moore, Bean, Birdyshaw, & Rycik, 1999) which advocates seven principles for adolescent literacy.

Another organization that has done a great deal to address the needs of adolescent readers is the Alliance for Excellent Education (AEE). The AEE's interest in adolescent

literacy emerged with the publication of *Every Child a Graduate: A Framework for an Excellent Education for All Middle and High School Students* (Joftus, 2002). This report highlights the fact that low-literacy levels affect more than a student's performance in his or her language arts class; low-literacy prevents students from doing well in their content area courses as well, and this places students in great risk of dropping out of school altogether. Exacerbating this situation is the fact that most high school teachers are not trained to teach reading despite the fact that research shows that students who receive intensive instruction and tutoring in reading will graduate from high school and pursue postsecondary education in higher numbers than those who do not receive such instruction. One of the four initiatives outlined in the report to address these issues was the formation of an Adolescent Literacy Initiative. The AEE created an advisory group to oversee the work of the initiative, and the group is comprised of such scholars as Donna Alvermann, Michael Kamil, Dorothy Strickland and Chris Tovani, just to mention a few. Under the leadership of this group, the AEE has released three important reports on adolescent literacy.

*The Literacy Coach: A Key to Teaching and Learning in Secondary Schools* (Sturtevant, 2004) introduces the concept of placing literacy coaches in secondary schools. The literacy coach's role might include working directly with struggling readers, but the main function of the coach is to oversee the development of the school's entire literacy program to include staff development issues. The coach works with content area teachers to help them utilize strategies to improve their students' abilities to read, write and succeed in the content area by combining literacy instruction with content area

curriculum requirements. Literacy coaches are crucial to meeting the needs of adolescent readers because, currently, content area teachers do not feel adequately prepared to meet students' literacy needs (Blintz, 1997; Ratekin, Simpson, Alvermann, & Dishner, 1998).

According to work done by Buehl (1999):

[h]igh school teachers as a group continue to feel isolated and hard pressed to bridge the span of abilities and learning demands of their classrooms. Preservice requirements for secondary reading have not been notably popular with beginning teachers, and inservice presentations have not necessarily been a satisfactory vehicle for supporting content teachers as they attempt to adjust the way they work with students. (p. 6)

Even middle and secondary language arts teachers are under prepared to work with struggling readers as most pre-service programs in secondary education focus on methods of teaching writing, literature and grammar, not on the teaching of reading (Irvin, Buehl, & Klemp, 2007). The AEE's recommendation for literacy coaches has been met with favor, and, in an unprecedented move, five professional organizations-- International Reading Association, National Council of Teachers of English, National Council of Teachers of Mathematics, National Science Teachers Association and National Council for the Social Studies--created a task force to design national standards for literacy coaches working with teachers across the curriculum (International Reading Association, 2005).

The AEE followed this publication with a comprehensive review of the literature on adolescent literacy by Michael Kamil in *Adolescents and Literacy: Reading for the 21<sup>st</sup> century* (2003). In the review, Kamil examines the reliable, empirical research that exists on how to improve the literacy skills of children in grades 4-12. Much of the

information is culled from four sources: the National Research Council report *Preventing Reading Difficulties in Young Children* (Snow, Burns & Griffin, 1998); the report of the National Reading Panel, *Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and its Implication for Reading Instruction* (National Institute of Child Health and Human Development, 2000); the *Handbook of Research on Reading* (Alvermann & Moore, 1991); and the RAND Reading Study Group's *Reading for Understanding: Toward a RAND Program in Reading Comprehension* (RAND, 2002).

The goal of the report is to compile the key findings of all of the best research available regarding adolescent literacy to offer the public and policymakers a more thorough understanding of what is known about the topic. Two important pieces of information emerged from this report. First, 1 out of 10 adolescents has serious difficulties identifying words (Curtis & Longo, 1999). These difficulties often result from problems associated with phonological aspects of word analysis. This problem becomes compounded in adolescence because adolescents abandon the process of trying to read the word and instead try to guess the word based on context clues. Hence, Kamil's report was one of the first to highlight the fact that some adolescents' literacy problems stem from much more basic difficulties; previously, comprehension skills alone were thought to be the primary culprit. The second item of import to emerge was that we actually know

and agree upon quite a bit with regard to adolescent literacy. On this finding, Meltzer (2002) writes:

I expected to find much more conflict about what we know. I expected to find hesitant recommendations and contradictory advice. Instead, I found resounding agreement across research communities about what is effective and what needs to be done (p. 106).

Kamil's report outlines findings about four general topics--the use of reading strategies, the prevention of reading difficulties, the components of effective reading instruction and reading comprehension--before discussing more specialized reading research in the area of English-language learners, computer-assisted instruction, teacher education and professional development. Finally, Kamil outlines an infrastructure for reading instruction in secondary schools.

Biancarosa and Snow (2004) built upon Kamil's report in *Reading Next: A Vision for Action and Research in Middle and High School Literacy* by delineating 15 elements aimed at improving adolescent literacy. The elements are divided into two sections: instructional improvements and the infrastructural improvements needed to support them. Each element is described and, in cases where proven practice exists, examples are included in the report. For instance, one of the elements listed as a critical component of a successful literacy program is that effective instructional principles must be embedded in content. A program that promotes this element is the Reading Apprenticeship Model (Jordan, Jensen, & Greenleaf, 2001) which is explained in detail.

The report continues with a call for educators and researchers to implement variations of the 15 key elements with a view toward not only improving student

achievement in reading but also improving the research base. Biancarosa and Snow (2004) introduce the concept of “planned variation” of elements whereby researchers study how certain factors interact when different elements of the 15 key components are combined with different populations of students. This approach would allow comparisons across programs and allow the value-added contribution of certain elements to be determined. In order for this to work, all programs would need to be measured on the same outcomes using common measures.

Biancarosa and Snow (2004) suggest 10 such outcomes, as well as the use of control groups and the use of specific reporting procedures. In conclusion, Biancarosa and Snow (2004) state that there is no “optimal mix” of the 15 elements. Choices of which elements to incorporate into a school’s program need to be matched to school and student need. Of the 15 elements, however, Biancarosa and Snow (2004) state the 3 elements of professional development and ongoing formative assessment and summative assessment should be the foundation of any program. The challenge issued by the report is to enhance achievement in adolescent literacy while at the same time extending the knowledge base in the field.

Since the year 2000, many other organizations have contributed to the call for improving adolescent literacy. The National Council of Teachers of English issued a position statement, *A Call to Action: What We Know about Adolescent Literacy and Ways to Support Teachers in Meeting Students’ Needs* (2004), and in 2005, the National Association of State Boards of Education, the National Association of Secondary School



Principals and the National Governors Association each issued lengthy reports addressing approaches to improving the literacy skills of adolescents.

### *1.1.2 Federal Support for Adolescent Literacy*

State and federal mandates have also had a large influence on adolescent literacy issues. The National Assessment of Educational Progress (NAEP) has provided an assessment of what America's students know and can do in various subject areas since 1969. It is the data gathered and analyzed by the National Center for Educational Statistics, the group responsible by law for carrying out the NAEP project, which has drawn attention to the shortcomings of our nation's attempts to address adolescent literacy. It is also data from this group that most of the aforementioned reports draw upon to support their point that adolescent literacy is not receiving the attention it requires. Aside from a period in the 1970's when teacher preparation courses entitled "Teaching Reading in the Content Areas" were instituted in most colleges of education, little seemed to be changing with regard to reform efforts in adolescent literacy until NCLB.

The NCLB Act heightens Title 1 accountability by requiring states to implement state-wide accountability systems covering all public schools and students. The accountability systems established by each state have to be based on challenging state standards in reading and mathematics, annual testing for all students and annual statewide progress objectives to ensure that all groups of students reach proficiency. Results of the testing must be disaggregated by socio-economic status, race, ethnicity, disability and limited English proficiency to guarantee that no group is left behind. School districts and schools that fail to make adequate yearly progress (AYP) toward proficiency goals are

subject to improvement, corrective action and restructuring measures to meet state standards.

In order to assist school districts' efforts for making adequate yearly progress, the NCLB Act includes the Reading First initiative to ensure that every child can read by the end of third grade by significantly increasing the Federal investment in scientifically-based reading instruction programs and awarding six-year grants to states. In turn, the states make competitive sub-grants to local communities and local recipients administer screening and diagnostic assessments to determine which students in grades K-3 are at risk of reading failure and provide professional development for teachers of grades K-3 in the essential components of reading instruction.

Another initiative introduced as a component of the NCLB Act was the Early Reading First program. This program awards competitive six-year grants to local education agencies to support early language, literacy and pre-reading development of preschool-age children, particularly those from low-income families. Recipients must use instructional strategies and professional development drawn from scientifically-based reading research to help young children gain the knowledge and skills they will need for optimal reading development in kindergarten and beyond (United States Department of Education, *Overview: Reading First*, 2006).

While most literacy researchers laud these two initiatives and the support that is being generated for emergent and early literacy, critics call for the same level of support for adolescent literacy. The 15-year-old ninth grader struggling with reading can no longer be ignored.

In late 2003, the AEE and almost two hundred organizations and individuals signed a letter to the President requesting him to address the needs of struggling adolescent readers in the fiscal year 2005 budget. Biancarosa and Snow's (2004) *Reading Next: A Vision for Action and Research in Middle and High School Literacy* sounded a call for reform as well. President Bush responded by proposing that \$100 million be spent on a new discretionary grant program entitled Striving Readers. In the final passing of the program, however, Congress appropriated only \$24.8 million.

The goal of the Striving Readers program is to increase the reading achievement levels of middle and high school-aged students in Title 1-eligible schools with significant numbers of students reading below grade level. The program supports new comprehensive reading initiatives or expansion of existing initiatives that improve the quality of literacy instruction across the curriculum, provide intensive literacy interventions to struggling adolescent readers and help to build a strong, scientific research base for identifying and replicating strategies that improve adolescent literacy skills (United States Department of Education, *Overview: Striving Readers*, 2006).

Another legislative initiative that has garnered much support from professional organizations concerned with adolescent literacy is the Pathways for All Students to Succeed (PASS) Act (S.921). The act is requesting that \$1 billion dollars be appropriated for literacy coaches to be placed in all middle and high schools at a ratio of 1 coach to every 20 students. The coaches would work with teachers to incorporate research-based literacy instruction in all academic classes, assess student progress, assist with diagnostic tests and provide school leadership with assistance in developing a school-based literacy

program. The PASS Act also provides for \$500 million to be spent on grant programs to turn around failing schools by helping them implement comprehensive reform initiatives (AEE, 2005).

The Graduation for All Act (H.R. 547) is another promising piece of legislature that has been proposed. The act includes the Improving Adolescent Literacy Initiative that would target secondary schools with the lowest graduation rates and provide funding for at least one literacy coach per school. These two bills, the Senate PASS Act and the House Graduation for All Act, along with the already approved Striving Readers Program, would begin to even the support in federal funding for literacy initiatives toward elementary and secondary students (AEE, 2005).

Hence, while the NCLB Act was passed in 2001, it seems that support for it at the secondary level was not forthcoming until Striving Readers was passed in 2005. In an attempt to build upon the Striving Readers Program, the Senate PASS Act and the House Graduation for All Act are also promising pieces of legislation. All three initiatives focus on professional development for teachers who work with struggling adolescent readers. It is becoming apparent to many that employing and retaining highly qualified teachers is of paramount importance.

### *1.1.3 The Push for Highly Qualified Teachers*

The single most important factor affecting students' academic achievement is an effective, qualified teacher (Hearing 106<sup>th</sup> Congress, 1<sup>st</sup> session, 1999; Hunt, Jr., 2003; Keegen, 2003). Ineffective teachers are detrimental to student academic achievement, and students who spend two consecutive years in the classrooms of ineffective teachers may

never recover to reach their full academic potential even if their ensuing teachers are effective (Sanders, 1998).

Although it has been two decades since the release of the seminal report *A Nation at Risk*, student academic achievement still remains a primary concern. The United States was the first nation to achieve universal elementary education, to expand the secondary system to include a diverse student population and to create an open competitive tertiary system of education. Yet, according to Peterson (2003), the United States still ranks significantly below other advanced industrial democracies and some developing countries in student academic achievement. The Koret Task Force stated that “[h]igher-quality teachers are key to improving schools, but the proper gauge of quality is classroom effectiveness” (Hunt, 2003, p.13). Since teachers are the dominant classroom input, classroom effectiveness means teacher effectiveness.

But what are the qualities of an effective teacher? “Over the last half of the century, researchers and school policymakers have sought to pin down precisely what teaching effectiveness is, which teachers have it, and how they got it” (Cuban, 1995, p. 230). Depending on the time period, the definition of teaching effectiveness changes: a century ago, teacher effectiveness would have been rated on the teacher’s personal characteristics, such as personality and deportment, rather than on performance in the classroom (Borich, 1992). Current definitions, though, center primarily on student outcomes, which are assumed to be based on teacher effectiveness.

Much of the research on teacher effectiveness focuses on observable teacher behaviors. According to Clark and Peterson (1986), however, teaching encompasses two

major domains. The first domain is teachers' thought processes, and the second domain is teachers' actions and their observable effects. Teachers' thought processes are teachers' cognitive activities that go on inside a teacher's head; they are not observable. Teachers' actions and their observable effects, on the other hand, include teacher actions, student actions and student achievement; they are observable and measurable. Traditional research on teaching has focused on the observable and easily measured teacher actions and their effects on student behavior and achievement. This body of research is often referred to as teacher effectiveness research. A basic assumption of this research is the belief that the relationship between a teacher's actions and the observable effects of those actions are linear and unidirectional; the teacher's actions affect student behavior which affects student achievement. Recently, however, researchers are beginning to view the direction of causality as circular whereby a teacher's actions affect student behavior which affects teacher behavior and ultimately student achievement. This new conception of causality, combined with advances in cognitive psychology and a broadening of research paradigms (Fang, 1995), has prompted researchers to become increasingly interested in the domain of teacher thought processes.

The potential connection between teacher effectiveness, in its newest incarnation of an effective teacher is a teacher who positively influences student outcomes, and student achievement is quickly being accepted as fact as many school districts nationwide have begun to implement incentive pay programs for teachers.

Denver, Colorado adopted such a plan in November of 2005. Houston, Texas followed suit in January the following year. The Houston plan allows teachers to earn

bonuses three ways. The first awards bonuses to teachers in schools rated acceptable or higher due to students' scores on the state's standardized test. The second bonus award ties teacher pay to student improvement on a standardized test that compares student performance to test norms nationwide. The third bonus is for math and language arts teachers whose students outperform other students district-wide on the Texas' mandated test ("Houston ties teacher pay to students' scores," 2006). Each of the three bonuses assumes a connection between teacher effectiveness and student performance.

The state of Florida unveiled a similar proposal, the "Effectiveness Compensation" plan, in February of 2006. Unlike the Denver and Houston plans, however, the Florida plan is the nation's first statewide merit pay plan. Ex-Florida Education Commissioner, John Winn, reasons, "If we are going to attract and keep the best and brightest teachers, then we must reward excellence in what matters most, student learning" (Kaczor, 2006). This type of reasoning weighs heavily on language arts teachers as most state-mandated testing is comprised of three separate tests: one in mathematics, one in writing, and one in reading.

Thus, the burden of student performance falls mainly upon language arts teachers. There has been, however, little research conducted to ascertain what it is exactly that makes a language arts teacher effective in light of this new determinant of student achievement. Similarly, little research has been conducted that takes into consideration this new concept of a circular causality in the relationship between a teacher's actions and the observable effects of those actions on student achievement by looking more closely at teacher thought processes and beliefs rather than teacher actions.

#### *1.1.4 Teacher Effectiveness and Teacher Efficacy*

One construct of teacher belief that has consistently been associated with the numerous qualities of an effective teacher, however, has been that of teacher efficacy. Tschannen-Moran and Woolfolk Hoy (2001) defined teacher efficacy as a teacher's "judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who may be difficult or unmotivated" (p. 283).

Teacher efficacy has been related to student outcomes such as motivation (Midgley, Feldlaufer & Eccles, 1989), achievement (Ashton & Webb, 1986) and students' own sense of efficacy (Anderson, Green & Loewen, 1988). Teacher efficacy has also been related to teacher behavior in the classroom. Allinder (1994) found that teachers with high efficacy beliefs plan more thoroughly and are more organized. Highly efficacious teachers have a willingness to try new strategies (Guskey, 1988), persist when teaching becomes difficult (Coladaraci, 1992) and work with struggling students longer (Gibson & Dembo, 1984). Teachers with high efficacy show greater commitment to teaching (Evans & Tribble, 1986), more enthusiasm for teaching (Allinder, 1994) and are more likely to stay in the profession of teaching (Burley, Hall, Villeme & Brockheimer, 1991).

The relationship between teacher efficacy and student achievement, however, is often indirect. Highly efficacious teachers behave in certain ways, and those behaviors result in increased student achievement. Ashton and Webb (1986) found that teachers with a high sense of efficacy have high expectations for all students, establish classroom



environments that encourage warm interpersonal relationships and promote strong academic work. They are more humanistic in their classroom management style (Woolfolk & Hoy, 1990), evoke trust from their students (DaCosta & Riordan, 1996) and favor student-centered classrooms (Czerniak & Schriver, 1994) as well as activity-based and experiential learning (Enochs, Scharman & Riggs, 1995). Furthermore, teachers with high efficacy are more likely to seek assistance from other educational professionals (DeMesuita & Drake, 1994) and promote parental involvement in schooling (Hoover, Dempsey, Bassler & Brissie, 1992).

The concept of teacher efficacy appears to be a promising construct in the research of teacher thought processes and teacher beliefs. In his research on teacher beliefs, Nespor (1987) wrote that there are two plausible lines of response to implications of the power of teacher belief systems. One of the lines of response is as follows:

The first would be to address the features of teaching as an occupation that make belief so important. One could, for example, routinize teaching to the extent that teachers could be taught recipe-like pedagogical methods, adherence to which could be closely monitored and regulated. That is, one could transform teaching into a set of well-defined tasks and thus reduce the role played by beliefs and shaping tasks. (p. 326)

The second plausible lines of response to implications of the power of teacher belief systems would be to try to change or shape teachers' beliefs rather than to "teacher-

proof” curricula so teacher beliefs do not come into play. In order to do this, however, teacher beliefs must be understood more explicitly. Nespor (1987) wrote:

In part this would mean helping teachers and prospective teachers become reflexive and self-conscious of their beliefs...However, this can result in the transformations of teachers’ beliefs and practices only if alternative or new beliefs are available to replace the old. This is the crux of the problem: we do not know very much about how beliefs come into being, how they are supported or weakened, how people are converted to them, and so on...[These] are questions that must be addressed. (p. 326)

## 1.2 Statement of the Problem

Many states, however, are choosing to routinize teaching as they move away from decisions about reading being made at the local level and, instead, talk about a "national reading curriculum.” This movement characterizes teachers as technocrats of the state who merely “deliver” a mandated curriculum (Allington, 2002). While this approach may sound a bit drastic to some, the public is being encouraged to agree with this movement. Consider the first line of a newspaper article entitled “Sticking with the ABCs of Reading:”

If a neurosurgeon were performing brain surgery, the last thing the patient would want would be for the doctor to get a sudden burst of creativity and deviate from established and accepted procedures. The same can be said for the teaching of reading. (Behrendt, 2004)

Holding teachers highly accountable while giving them little autonomy lessens the likelihood that teachers will accept professional responsibility for providing effective instruction (Bean & Harper, 2004). An example of the ineffectiveness of this approach can be found in a recent assessment conducted of the Read 180 program, a program

designed to help struggling readers make a turn around in their reading ability. According to the developer of the program, Scholastic, 1 out of every 40 adolescent students in the United States has used this program. The report of the program's effectiveness in one school district in the southwest concluded that "only a fraction of the 128 Read 180 classrooms in the district were faithfully following the program's model...educators...failed to see that the program's highly prescriptive lessons were carried out the way they should be" (Tobin, 2006). The report attributes this to the fact that the district recruited "teachers who wanted no part of the program" (Tobin, 2006). Holding teachers highly accountable while giving them little autonomy also contradicts research indicating that it is autonomous, expert teachers who increase student gains in learning.

In the very first research done with regard to teacher efficacy and student achievement in reading, Armor et al. (1979) concluded that a good reading program:

...features school autonomy and teacher flexibility. Because principals seem to be able to identify effective practices, because teachers need the freedom to adapt and the training for individualization, and because no prefabricated reading system or strategy can provide a "quick fix" to the complex problem of reading instruction, a good measure of decision making authority at the local school level seems warranted. (p. vi)

Biancarossa and Snow (2004) echoed this same sentiment 25 years later in their report *Reading Next: A Vision for Action and Research in Middle and High School Literacy* in which they delineate 15 elements aimed at improving adolescent literacy. They state that there is no "optimal mix" of the 15 elements. Choices of which elements to incorporate into a school's program need to be matched to school and student need.

Darling-Hammond (1999), too, supports the idea of investing in teachers rather than investing in teacher-proof curricula when she found achievement levels in reading rose in states that invested heavily in teacher development while there was little progress evident in states that invested heavily in testing and curriculum standards.

This claim reiterates the findings of the National Commission of Teaching and America's Future (1997):

Investment in teacher development produced far greater student achievement gains than [other] investments...spending on teacher development swamped other variables as the most productive investment for schools. (p. 9)

Woolfolk and Hoy (1990) stated that although “researchers have found few consistent relationships between characteristics of teachers and the behaviors or learning of students, teachers’ sense of efficacy is an exception to this general rule” (p. 81). Because of this, individual teacher efficacy is a teacher belief variable worthy of continued study if we hope to raise the reading achievement levels of all students through the professional development of teachers.

As states hold teachers more and more accountable for student achievement, districts must find ways to provide professional development for teachers in a strategic manner. A more complete understanding of the impact of teacher efficacy in the areas of instructional strategies, classroom management and student engagement on student achievement in reading would provide much needed insight into the types of targeted professional development that would assist teachers in promoting highest student achievement in reading for all students.

### 1.3 Purpose of the Study

The purpose of this study was to investigate the effects of teacher efficacy for instructional strategies, classroom management and student engagement on ninth and tenth grade students' reading achievement. Teachers reporting high efficacy were hypothesized to positively impact students' reading gains. Student contextual variables of prior student achievement, socioeconomic status, ethnicity and grade were controlled for in the study.

### 1.4 Research Questions

While taking student characteristics (prior student achievement, SES, grade and ethnicity) into consideration:

1. What is the influence of teacher efficacy regarding instructional strategies on student reading achievement gains?
2. What is the influence of teacher efficacy regarding classroom management on student reading achievement gains?
3. What is the influence of teacher efficacy regarding student motivation on student reading achievement gains?
4. What is the influence of total teacher efficacy (teacher efficacy regarding instructional strategies, classroom management and student motivation combined) on student reading achievement gains?

### 1.5 Methodology

In order to investigate these questions, the statistical analysis was not stepwise multiple regressions, the statistical procedure used in most studies of this type, but rather

hierarchical linear modeling. This research investigates the degree to which teacher efficacy (a class-level factor) in three areas, instructional strategies, classroom management and student engagement, relates to student achievement in reading (an individual-level outcome). Because student achievement occurs at the individual level and teacher efficacy influences students at the classroom level, the unit of analysis becomes problematic. Conventional methods require single-level analysis that leads to the conceptual and empirical problems associated with examining student-level and teacher-level variables.

Multilevel modeling, on the other hand, avoids misestimation of standard errors, aggregation bias and heterogeneity of regression problems that may affect the results of ordinary least squares regression analysis of data in which one or more individual-level characteristics are aggregated to the group level. Hierarchical linear modeling (HLM) is a multilevel modeling technique intended for nested data (Bryk & Raudenbush, 2002). HLM takes into account the interdependence of individual measures collected within the same organizational unit, the classroom in this case.

In addition, HLM partitions the variance in a dependent variable into its within- and between-classroom components. Therefore, it is possible to model the effect of teacher efficacy on only the portion of the variance in student achievement occurring between classrooms while modeling the influence of students' SES, grade, ethnicity and prior achievement without aggregating these variables at the classroom level. In other words, it is possible to predict values of the dependent variable based on a function of predictor variables at more than one level (Luke, 2004).

## 1.6 Conceptual Framework

The grounding for this study is Albert Bandura's (1977, 1997) social cognitive theory, which assumes that people are capable of human agency and that human agency operates in a process called triadic reciprocal causation. The triad stems from three inter-related forces: behavior, environmental influences and internal personal factors resulting from affective, cognitive and biological processes. Collaboratively, these three forces impact what people believe about themselves as well as the actions they will take.

Central to Bandura's theory is his concept of self-efficacy. In his seminal work, "Self-efficacy: toward a unifying theory of behavioral change," Bandura (1977) defined self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Consistent with the general formulation of the concept of self-efficacy, Tschannen-Moran and Woolfolk Hoy (2001) defined teacher efficacy as a teacher's "judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who may be difficult or unmotivated" (p. 283). In the field of educational research, however, application of Bandura's concept of efficacy has been intermingled with Rotter's locus of control theory (1966) and Weiner's (1979) attribution theory through the instruments used to measure teacher efficacy.

Rotter's theory (1966) also focuses on human agency, but he posits that human agency was determined by whether an individual possessed an internal locus of control, believing that the ability to influence a situation lay within them, or if they possessed an

external locus of control, believing that the power to change a situation lay not within them but within the environment.

The fundamental assumption of Weiner's attribution theory (1979) is that stability, controllability and locus of causality together have power in predicting important behavioral outcomes, emotions and motivation. Rotter's locus of control and Weiner's attributional theory are similar to the extent that both address the connection between attributed cause and outcomes. Bandura's self-efficacy theory (1977, 1997) expands the response-outcome expectancies to include efficacy expectations. Efficacy expectations dictate that expectancies will have no effect on behavior unless the individual believes that she/he can successfully execute or carry out the behavior required to produce the outcome. Despite the fact that the theoretical formulation of the construct of teacher efficacy has become somewhat muddled through psychometric attempts to measure the concept, teacher efficacy is a variable worthy of study.

### 1.7 Significance of the Study

Teacher efficacy is both context- and subject-matter specific (Pintrich & Schunk, 1996), and efficacy is not uniform across the many different types of tasks that teachers are asked to perform (Bandura, 1997). In light of this, it seems that global measures of teacher efficacy, such as the commonly used Gibson and Dembo Teacher Efficacy Scale (1984), may obfuscate important information about teacher efficacy. Additionally, the Teacher Efficacy Scale (TES) has been demonstrated to have problems in both reliability of its subscales (Henson, Kogan & Vacha-Haase, 2001) as well as its factor structure (Guskey & Passaro, 1994). The meaning of research findings based on this measure is



unclear (Deemer & Minke, 1999). The present study utilized the Ohio State Teacher Efficacy Scale (OSTES) rather than the Gibson and Dembo Teacher Efficacy Scale (1984). The OSTES divides efficacy into three factors: efficacy for instructional strategies, efficacy for classroom management and efficacy for student motivation, allowing the researcher to investigate teachers' level of efficacy with regard to the various acts they are asked to perform during a school day (Tschannen-Moran & Woolfolk Hoy, 2001).

Furthermore, the current body of research focuses on teachers of basic skills classes or elementary school generalists. The sample for the present study was teachers of English / Language Arts in a multitude of public school settings at the secondary school level. Secondary grade levels were specifically chosen because, according to the statewide results of the 2005 Florida Comprehensive Assessment Test (FCAT), 71% of fourth graders, 44% of eighth graders and 32 % of tenth graders were proficient readers. At the proficient level, students are able to make inferences more easily as well as understand character motive and identify themes and literary devices used by the author (Dunn, 2006). The steady decline in the percentage of students reading at a proficient level in the secondary years is especially troublesome.

While there has been much examination of teacher efficacy and its relationship to various aspects of schooling, research that attempts to uncover the extent to which teacher efficacy relates to student achievement in reading is sparse. To date, there have only been five studies directly examining this connection, and most of these studies included classrooms of teachers of grade 6 or lower. Tracz and Gibson (1986) examined

teachers in grades 4 – 6; Anderson, Green and Loewen (1988) focused on teachers of grades 3 and 6; and Borton's (1991) sample included teachers of grades 3 and 4. Armor et al. (1979) carried out studies with middle level teachers in grade 6 and 7 classrooms while Ashton and Webb (1986) conducted the only study at the high school level with teachers of grades 9 – 11. Clearly, more research is needed in the area of student reading achievement at the secondary school level. This study focused on teachers and students in grades 9 and 10.

Finally, it is apparent that the statistics for student achievement in reading become extremely disparate when scores are disaggregated by groups. Only 15% of low-income eighth graders read at grade level, and the achievement gaps between White and Black students, White and Hispanic students and those who were not eligible for free or reduced-price lunch and those who were eligible have shown no significant changes between 1992 and 2005 (Perie, Grigg, & Donahue, 2005). Unlike other studies of this nature, the contextual variables of socio-economic status, grade and ethnicity were taken into consideration in the statistical analysis.

### 1.8 Basic Assumptions

One assumption of the present study is that the FCAT is a valid and reliable measure of reading achievement. The only technical information about the FCAT is provided in publications authored by the Florida Department of Education (FLDOE) or by agencies working under the auspices of the FLDOE. It has been assumed that these reports are accurate.

Another postulate is that the construct of efficacy as developed by Bandura's social cognitive theory (1977, 1997) adequately captures the cognitive processes that serve to motivate individuals. Bandura believes that cognitive processes play a part in the course of behavioral change in that an individual's cognitive representations of anticipated outcomes can motivate behavior. He hypothesizes that these cognitive representations produce the expectation in an individual that by behaving in a certain way, he or she can create positive outcomes and preclude difficulties. His theory of efficacy also stipulates that goal setting and self-evaluation work as cognitively based motivators and that all of these motivating factors have an effect on both the initiation and the persistence of individuals to reach a certain level of performance. Tschannen-Moran and Woolfolk Hoy (2001) applied Bandura's construct of efficacy to the domain of teaching and created the OSTES to measure teacher efficacy. This study also assumes that the OSTES accurately reflects and assesses Bandura's construct of efficacy.

The fact that there is a relationship between student academic achievement in reading and teachers' ratings of their own efficacy as measured by the OSTES is another assumption of the study. It was also assumed that there is a relationship between student academic achievement in reading and teachers' own ratings in the individual subscale categories of teacher efficacy for instructional strategies, classroom management and student motivation.

The idea that teachers' perceptions of their self-efficacy were accurate and that teacher behavior was consistent with their reported level of efficacy was another supposition. In other words, teachers who perceived themselves as having high efficacy

were presumed to behave in efficacious ways in their interaction with the students in their classrooms.

### 1.9 Limitations

There were several limitations to this study. The first of these was the need to use criterion sampling (Creswell, 1998) in order to involve teacher participants who taught students participating in the 2007 administration of the FCAT. The researcher attended district-wide meetings of English department chairs to explain the study and request that at one of their school's English department meetings, the chairs ask their ninth and tenth grade teachers to volunteer for the study. Each department chair was given a packet containing all of the materials needed to administer the survey as well as a script to read to their teachers.

Because the study was also limited to teachers of and students in regular and honors English Language Arts classes in grades 9 and 10 who attended schools in two different school districts on the west coast of Florida, the findings may not be generalizable to teachers of and students in other grade levels and courses nationwide.

Furthermore, the sample of students in this study included students in both regular English classes and honors English classes. The influence of teacher efficacy may have been easier to detect if research focused on students in pure reading classes rather than students enrolled in the more traditional English classes where all five strands of the language arts curriculum as defined by the National Council of Teachers of English - reading, writing, speaking, listening and viewing - are deemed equally important. Teachers of traditional English classes integrate all of these facets into instruction rather

than concentrating solely on improving student reading. Students in pure reading classes would have mirrored Ashton and Webb's (1986) work with basic skills students, however, and the goal of this research was to investigate the influence of teacher efficacy on a different student population. Reading classes in the two districts included in the study are comprised of students who have failed standardized reading tests.

An additional condition that limits the generalizability of the study is the fact that the dependent variable in the study was measured with one single item, the developmental scale scores on the reading portion of the FCAT. It would be preferable to measure reading achievement gain with several instruments that could be tested for internal validity and consistency. Thus, the findings of this study may not generalize to other measures of student achievement in reading. Furthermore, the researcher was not involved in the administration of the FCAT; hence, it can only be presumed that the test was administered under the guidelines prescribed by the Florida Department of Education.

Another limitation of the study is that the construct of teacher efficacy cannot be measured directly. As a result, this study relied on self-report data. Anastasi (1982) suggests that self-report measures can be subject to inaccurate evaluations by respondents. Incidents of this, however, have not been a topic of concern in the literature regarding the measurement of teacher efficacy. Furthermore, on a variable that is internal to the respondent, self-report data are considered more reliable than third-party observations (Starnaman & Miller, 1992).

It should also be noted that the surveys were completed by teachers three weeks after the 2007 FCAT administration and one week prior to Spring Break. On a positive note, teachers have a strong knowledge of their students by this point in the academic year. Because of this, teachers may have been able to provide a more adequate depiction of their sense of efficacy in delivering instruction to their students than at an earlier point in the school year. Teacher fatigue, however, is also high directly after large-scale testing takes place and when Spring Break is on the horizon. It is hoped that the teachers' knowledge of their students outweighed any fatigue they might have been feeling with regard to the time of survey administration and the accuracy of teachers' self-assessment of efficacy.

Finally, two statistical limitations must be mentioned. First, only a portion of the variance in reading achievement gain was accounted for by the variables examined in each model. It is possible that teacher, student, school and district variables of importance are missing from the models. The exclusion of such variables can lead to model misspecification or biased coefficients resulting in potentially misleading findings. The researcher attempted to avoid model misspecification by including the contextual variables of race, grade, SES and by measuring for reading achievement gain from 2006 to 2007 rather than overall reading achievement. Another statistical limitation results from the correlational design of the study. Such a design does not allow causal conclusions to be made and, thus, cannot be used to guide policies or programs related to ninth and tenth grade student achievement in reading.

## 1.10 Definition of Terms

The following terms used in the present study are defined alphabetically:

**Ethnicity.** The ethnic affiliation of a person as reported by the school district to the testing agency: African-American, Caucasian, Hispanic, or Other.

**Gender.** The sex of a person as reported by the school district to the testing agency: male or female.

**Hierarchical linear modeling (HLM).** Also commonly referred to as multilevel modeling, HLM is an analytic technique that is useful to examine data that are nested within one another, such as students within classrooms. HLM controls for the non-independence of observations that occurs due to this nesting as individuals who belong to a group are likely to be similar to one another, resulting in correlated data. HLM also allows for the examination of the variability within and between individuals and groups as well as their interactions (Hox, 2002).

**Influence.** According to *The Merriam-Webster Dictionary* (Mish et al., 2004), influence is defined as “the power or capacity of causing an effect in indirect or intangible ways” (p. 372).

**Intraclass correlation coefficient (ICC).** The intraclass correlation coefficient represents “the proportion of variance in a dependent variable that is between groups (i.e., Level-2 units)” (Raudenbush & Bryk, 2002, p. 36).

**Ohio State Teacher Efficacy Scale (OSTES).** A teacher efficacy measurement device recently developed by Tschannen-Moran and Woolfolk Hoy (2001). The OSTES is a 24-

item Likert-type instrument with a nine-point scale. The scale anchors at 1-Nothing, 2-Very Little, 5-Some Influence, 7-Quite A Bit and 9-A Great Deal.

**Statistical reliability.** Reliability refers to the extent to which an experiment, test, or any measuring procedure yields the same result on repeated trials.

**Statistical validity.** Validity refers the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure.

**Student reading achievement.** The variable was created by subtracting the 2006 FCAT developmental scale scores (DSS) from the 2007 FCAT DSSs. Developmental scale scores can range between 86 points and 3008 points. Creating this variable allowed the researcher to control for student prior achievement.

**Socio-economic status.** This is an economic designation assigned to students determined by a student's participation in the free and reduced lunch program. For purposes of this study, SES is divided into two categories: those students who are eligible for the free or reduced lunch program and those students who are not eligible for free or reduced lunch. According the website (School District of Hillsborough County, 2006) that explains eligibility guidelines, the National School Lunch and Breakfast program uses United States Department of Agriculture Income Eligibility Guidelines which are adjusted annually for inflation and are used when approving applications to receive free or reduced lunch. The free lunch guidelines are 130% of Federal poverty guidelines, and the reduced lunch guidelines are 185% of poverty guidelines.

**Teacher efficacy.** Teacher efficacy is a teacher's "judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students



who may be difficult or unmotivated” (Tschannen-Moran & Woolfolk Hoy, 2001, p. 283).

**Teacher efficacy for classroom management.** “Teachers’ beliefs in their capabilities to organize and execute the courses of action required to maintain classroom order” (Brouwers & Tomic, 2000).

**Teacher efficacy for instructional strategies.** Teachers’ beliefs in their capabilities to plan lessons, explain concepts, question and respond to student questions, employ alternative teaching strategies and assess student comprehension.

**Teacher efficacy for student engagement.** Teachers’ beliefs in their capabilities to promote motivated and active learning in students.

#### 1.11 Organization of the Dissertation

This dissertation is divided into five chapters. The first chapter provides an introduction to the study and outlines the purpose of the study as well as the conceptual framework of social cognitive theory, the theory upon which the study is built. The research questions are enumerated and the methodology used to address the questions is discussed. Terms used throughout the study are defined, and assumptions and limitations of the study are addressed.

Chapter two reviews the development of literacy skills beyond emergent literacy and current trends in adolescent literacy. Albert Bandura’s social cognitive theory, a theory that has guided much of the research in teacher efficacy, is discussed along with various methods of measuring the construct of efficacy and some of the problems associated with measuring teacher efficacy. Chapter two also reviews the literature

concerning the relationship between teacher efficacy and various aspects of school in general as well as the relationship between teacher efficacy and student achievement in reading specifically. Finally, since the OSTES, the instrument used to measure teacher efficacy in this study, divides the measure of efficacy into three sub-categories, teacher efficacy for instructional strategies, classroom management and student motivation, the literature for each sub-category is reviewed as well.

Chapter 3 describes the methodology used to answer the research questions. It includes a description of the participants in the study and the data collection process. The instruments are introduced, and the reliability and validity of each instrument is addressed. The statistical analysis of the data is outlined, and the hierarchical linear models are described. This is followed by a discussion of how results will be interpreted.

Chapter 4 includes the data and analysis of the study. Descriptive information is provided for both student and teacher participants. The confirmatory factor analysis is discussed in detail as are the hierarchical linear models used to answer the research questions. The findings for each research questions are also enumerated.

Chapter 5 provides background information for the study and re-introduces the purpose of the study as well as the research questions. There is a discussion of the implications of the findings as well as a section addressing directions for future research.

## 1.12 Summary

The reading scores of America's middle school and high school students have shown no improvement at all between the years of 1971 and 2004 (NCES, 2004). According to the latest National Assessment of Educational Progress report, more than 8

million of America's middle and high school students are struggling readers (NCES, 2003). When scores are disaggregated by groups, the statistics are even more disturbing. Only 15% of low-income eighth graders read at grade level, and the achievement gaps between White and Black students, White and Hispanic students and those who were not eligible for free or reduced-price lunch and those who were have shown no significant changes between 1992 and 2005 (Perie, Grigg, & Donahue, 2005). The fact that many adolescents struggle with reading is an issue that many acknowledged; the fact that little was being done to rectify the situation was largely unchallenged until the late 1990's. At this time, numerous professional organizations sounded a clarion call for reform with regard to adolescent literacy. The call was heard, and the federal government appropriated funding for a new discretionary grant program entitled Striving Readers. Other promising pieces of legislation, the Pathways for All Students to Succeed (PASS) and the Graduation for All Act, followed. Finally, adolescent literacy is receiving the support it needs both in research and in funding.

At the same time, the standards movement has refocused the nation's attention on the issue of employing and retaining highly qualified teachers. The interest in this goal has caused the public and researchers alike to ponder what it is that makes a teacher highly qualified, or effective. Research in teacher effectiveness is also undergoing a paradigm shift as researchers are no longer simply focusing on observable teacher actions and, instead, are turning their research efforts to understanding teacher beliefs. It is becoming apparent that many reform initiatives will not be successful if teacher beliefs are not taken into consideration. One construct of teacher beliefs that holds much promise

is the concept of teacher efficacy which is grounded in Albert Bandura's socio-cognitive theory. Teacher efficacy has been related to many student outcomes, including student achievement.

While many exciting strides have been made to improve adolescent literacy, we are currently at a crossroads in the journey. One path leads to "routiniz[ing] teaching to the extent that teachers could be taught recipe-like pedagogical methods, adherence to which could be closely monitored and regulated" (Nespor, 1987, p. 326); the other path leads to the study and understanding of teacher beliefs in an effort to improve student achievement through meaningful professional development.

The NCLB Act has challenged educational researchers to go beyond student-level variables to search for school-level variables that increase student achievement. Researchers have found a positive relationship between student achievement and three kinds of efficacy beliefs: students' self-efficacy beliefs (Pajares, 1996), teachers' self-efficacy beliefs (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) and teachers' collective efficacy beliefs about the school (Goddard, Hoy, & Woolfolk Hoy, 2000). This study focused on individual teacher efficacy and student achievement because individual teacher efficacy is a school property amenable to change through the professional development opportunities that new legislation in adolescent literacy will afford to school districts.

## CHAPTER 2

### REVIEW OF THE LITERATURE

The purpose of this study was to investigate the effects of teacher efficacy for instructional strategies, classroom management and student engagement on ninth and tenth grade students' reading achievement. Teachers with high efficacy were hypothesized to positively impact students' reading gains. Student contextual variables of prior student achievement, socioeconomic status, ethnicity and grade were controlled for in the study.

This chapter first reviews important aspects of adolescent literacy: the development of literacy skills beyond emergent literacy and current trends in the study of adolescent literacy. The broad concept of teacher beliefs is then discussed prior to an in-depth explanation of Albert Bandura's social cognitive theory, a theory that has guided much of the research in one concept of teacher beliefs called teacher efficacy. Various methods of measuring the construct of efficacy will be reviewed as well as some of the problems associated with measuring teacher efficacy. This chapter then outlines the literature concerning the relationship between teacher efficacy and various aspects of school in general as well as the relationship between teacher efficacy and student achievement in reading specifically. Finally, since the OSTES, the instrument used to measure teacher efficacy in this study, divides the measure of efficacy into three sub-

categories, teacher efficacy for instructional strategies, classroom management and student motivation, the literature for each sub-category will be reviewed.

## 2.1 Adolescent Literacy

While researchers in the field of adolescent literacy concur that the No Child Left Behind (NCLB) Act's goal of having all children reading at grade level by third grade is meritorious, many in the field feel that NCLB has, indeed, left many children behind.

Hock and Deshler (2003) called attention to this fact in "*No Child*" *Leaves Behind Teen Reading Proficiency* when they concluded:

Although NCLB holds great promise for reforming U.S. schools, its effects will not be realized for many years. Likewise, the effects of the significant work done on early reading during the past decade by the National Institute of Child Health and Human Development will not affect millions of adolescents who are no longer in primary grades. The reality is that 15-year-olds who struggle with reading pose different challenges than those 5-year-old beginning readers pose. Solutions relevant to adolescent development and appropriate implementation within high school settings are desperately needed. (p. 35)

Literacy advocates now realize that the field of adolescent literacy is distinct from emergent and developing literacy and that the needs of adolescent readers must be met.

### 2.1.1 *The Development of Adolescent Literacy*

Since the 1920's, non-educators and educators alike who work with adolescents have recognized that older learners have literacy needs (Moore, Readance & Rickelman, 1983). Works such as Rudolf Flesch's controversial 1955 publication of *Why Johnny Can't Read: And What You Can Do About It* and Jean Chall's 1967 *Learning to Read: The Great Debate* ignited intense debate regarding the causes of what some termed as a

national literacy crisis (Henriquez, 2005). Chall followed this publication with *Stages of Reading Development* (1983) in which she outlined five stages of reading development.

According to Chall (1983), the pre-reading stage encompasses the time period when children are developing their oral language skills. During elementary school, most students experience the first three stages of reading. Initially, they learn to associate letters with sounds and to sound out words through decoding. During grades 2 and 3, children generally continue to increase their decoding skills, improve in their fluency and begin to develop various comprehension strategies. After third grade, the purpose of reading shifts; students are no longer learning to read but rather reading to learn.

From fourth grade on, students begin to read texts for the purpose of gathering information. They move beyond reading storybooks and basic readers and begin to encounter a larger variety of texts ranging from textbooks, biographies, laboratory instructions to mathematical word problems. Students must expand their vocabularies as well as their background and world knowledge. In addition, they must develop the comprehension and retention strategies necessary to obtain information from the texts. Ironically, it is also at this point in a child's education that most schools stop providing reading instruction. Many students, even those who were successful readers in the early grades, find this shift in the purpose of reading combined with the lack of support difficult and experience what Snow (2002) has coined *the fourth-grade slump*. These students become discouraged and begin to do less well academically (Kos, 1991; McCray, Vaughn, & Neal, 2001). As their motivation and confidence wane, they read less and less, falling further behind. Hence, good readers continue to improve as readers,

and poor readers struggle even more. Stanovich (1986) referred to this as *the Matthew effect*. For students who make this transition in reading purpose with ease, stage three continues until the end of the middle school years.

In high school and college, students who continue to develop as readers experience stages four and five where the language and cognitive demands placed upon them increase. In stage four, adolescents begin to analyze texts critically and to understand multiple points of view. During stage five, readers learn to take into consideration a significant amount of information of varying viewpoints and to construct their own knowledge based upon their analysis and synthesis of the information.

Reading scholar Richard Vacca recounts his own struggle with reading at stage five in his forward to *Reconceptualizing the Literacies in Adolescents' Lives* (Alvermann, Hinchman, Moore, Phelps & Waff, 2006) when he writes about his days as a doctoral student:

Soon into my doctoral program, I found myself struggling with learning – and doubting myself...Although I could read a best selling novel with ease in 5-6 hours, I found myself taking similar amounts of time to read chapter assignments from my statistics and methods textbooks – only to come up short in understanding what I had just read. I soon eschewed reading these “incomprehensible” texts altogether...On more than one occasion, I thought about quitting and returning to classroom teaching where I had been successful. (viii)

Despite the fact that most educators acknowledge that the development of strong literacy skills is a continual process, it is still commonly thought that reading instruction should be completed by grade 3 or 4 (Sturtevant, Boyd, Brozo, Hinchman, Moore &



Alvermann, 2006). And, according to Vacca and Alvermann (1998), most public discussions about reading instruction have neglected the needs of adolescents.

Since the publication of *Why Johnny Can't Read: And What You Can Do About It* (Flesch, 1955), Congress has legislated that substantial funds be spent on solving our literacy crisis. These funds have been earmarked for such initiatives as Title I for the Elementary and Secondary Education Act (now known as the No Child Left Behind Act, or NCLB), Reading First and Head Start. Each of these initiatives has emergent literacy as its focus and is geared toward getting all children reading on grade level by third grade. During the 2006 fiscal year, one billion dollars was spent on the Reading First Initiative while only 29.7 million was spent on the Striving Readers Program, a program authorized in 2005 to raise the reading achievement levels of middle and high school-aged students in Title I-eligible schools. During the 2003-2004 school year, Reading First spent \$72.00 per student in grades K-3 while the Striving Readers program only spent \$0.13 per student in grades 6-12 (Alliance for Excellent Education [AEE], 2006). While the goal of having all students reading on grade level by third grade is an important national priority since early reading skills are a strong predictor of long-term academic success, policymakers have disregarded the needs of struggling adolescent readers for too long.

### *2.1.2 Current Trends in the Study of Adolescent Literacy*

Inequities in our nation's priorities are illustrated clearly by the fact that while fourth grade reading scores have continued to improve over the years, reflecting the nation's focus on early literacy instruction, the reading scores of America's middle

school and high school students have shown no improvement at all between the years of 1971 and 2004 (National Center for Educational Statistics, 2004). As a matter of fact, the average reading scores for 9-year-old students were at an all time high in 2004; scores for 13-year-old students have risen only three points since 1975, and scores for 17-year-old students have dropped five points since 1992 (Perie, Moran, & Lutkus, 2005).

According to the latest National Assessment of Educational Progress, also known as “the nation’s report card” or NAEP, more than 8 million of America’s middle and high school students are struggling readers. Two-thirds of all eighth graders read below grade level, and of those, more than half score below what the U.S. Department of Education categorizes as its most basic level (NCES, 2003). At this level, students understand texts at a literal level and are able to make some interpretations. They are able to identify the main idea, extend some of the information via inferences, relate the text to personal experiences and draw conclusions (Perie, Moran, & Lutkus, 2005).

At the proficient level, students are able to make inferences more easily as well as understand character motive and identify themes and literary devices used by the author. Students categorized as advanced should be able to make complex summaries and theme statements, describe the interaction of literary devices and discuss the effects of those literary devices as well as be able to evaluate the author’s style and articulate their own reaction to the style (Perie, Moran, & Lutkus, 2005).

According to the 2005 report, 29% of the students fell below the basic level, 42% fell within the basic level, 26% were categorized as proficient and only 3% achieved at the criteria necessary for the advanced level (Perie, Grigg, & Donahue, 2005). The

statistics become even more disturbing when they are disaggregated by groups. Only 15% of low-income eighth graders read at grade level, and the achievement gaps between White and Black students, White and Hispanic students and those who were not eligible for free or reduced-price lunch and those who were eligible have shown no significant changes between 1992 and 2005 (Perie, Grigg, & Donahue, 2005). A policy brief published by the AEE opines that “it seems no coincidence that the national dropout rate closely mirrors the percentage of students reading at ‘below basic’ levels...”(AEE, 2006, p. 2). Low literacy levels affect learning in all subject areas (AAE, 2002), and poor readers who struggle in courses that require a great deal of reading are discouraged from taking more challenging courses (Au, 2000). It is not that surprising that more than 6,000 students drop out of high school every day (AEE, 2003). One of the reasons most commonly cited for dropping out is that students do not have the literacy skills necessary to be successful (Kamil, 2003; Snow & Biancarosa, 2003).

Struggling readers with college aspirations continue to confront difficulties in their post-secondary education. Research by ACT, *Reading Between the Lines: What the ACT Reveals about College Readiness in Reading* (2006), reports that only 51% of the high school graduates who took the ACT are prepared to handle the reading requirements of typical, first-year college coursework. Eleven percent of incoming university freshmen must enroll in remedial reading classes before they can even begin to take credit-earning courses (NCES, 2003), and out of that 11%, 70% do not attain a college degree within eight years of enrollment (Adelman, 2004).

For poor readers entering the workforce after high school graduation, frustration abounds as well. Many job-applicants must take an employer-administered test prior to being hired. In a survey published by the Center for Workforce Preparation (2002), 38% of job applicants lacked the reading skills necessary in the jobs for which they had applied. This percentage has doubled in four years due to the fact that reading requirements for many jobs have increased rapidly. In 2001, 80% of businesses reported that they had a serious shortage of candidates qualified for empty positions and cited poor reading skills as the primary reason (National Center of Manufacturers, Andersen, & Center for Workforce Success, 2001). It is abundantly clear that many adolescents are not being adequately prepared for the literacy demands of post-secondary education or of the workforce.

### *2.1.3 Summary of Adolescent Literacy*

Literacy advocates now realize that the field of adolescent literacy is distinct from emergent and developing literacy and that the needs of adolescent readers must be met. According to Jeanne Chall, adolescents who are good readers continue to develop their literacy skills in the years following elementary school (1983). Inequities in the attention and funding targeted for emergent literacy and adolescent literacy are becoming apparent, however, as fourth grade reading scores have continued to improve over the years while the scores of middle school and high school students have stayed the same or declined. Also distressing is the fact that the achievement gap among White, Black and Hispanic students has experienced no significant narrowing, nor has the gap between advantaged and disadvantaged students lessened (Perie, Moran, & Lutkus, 2005).

Low literacy levels for adolescent students increase the likelihood that poor adolescent readers will abandon their secondary education, experience difficulty in post-secondary education, if they pursue one at all, and will lack the reading skills necessary in today's workforce when they seek employment (Adelman, 2004; Center for Workforce Preparation, 2002). In an effort to seek ways to improve adolescent literacy, educational researchers are now looking beyond student-level variables to school-level variables that may increase student achievement (Ballou, Sanders & Wright, 2003).

## 2.2 Teacher Beliefs

One such variable is the classroom teacher, and a promising area of study with regard to the classroom teacher is teacher beliefs. According to Clark and Peterson (1986), teaching encompasses two major domains. The first domain is teacher thought processes and beliefs, and the second domain is teachers' actions and their observable effects. Teachers' thought processes and beliefs are teachers' cognitive activities that go on inside a teacher's head; they are not observable. The second domain of teachers' actions and their observable effects include teacher actions, student actions and student achievement; they are observable and measurable. Traditional research on teaching has focused on the observable and easily measured teacher actions and their effects on student behavior and achievement. This body of research is often referred to as teacher effectiveness research. A basic assumption of this research is the belief that the relationship between a teacher's actions and the observable effects of those actions are linear and unidirectional; the teacher's actions affect student behavior which affects student achievement. Recently, however, researchers are beginning to view the direction

of causality as cyclical or circular whereby a teacher's actions affect student behavior which affects teacher behavior and ultimately student achievement. This new conception of causality, combined with advances in cognitive psychology and a broadening of research paradigms (Fang, 1995), has prompted researchers to become increasingly curious about the domain of teacher thought processes.

Interest in teacher thought processes, while a latecomer to the research arena, dates back to Philip Jackson's (1968) *Life in the Classroom*. In this work, Jackson attempted to describe the mental thought processes that underlie teacher behavior. At the same time, Milton Rokeach (1968) published *Beliefs, Attitudes and Values: A Theory of Organization and Change* in which he defined a belief system as "having represented within it, in some organized psychological but not necessarily logical form, each and every one of a person's countless beliefs about physical and social reality" (p.2). His theory rested upon three assumptions: beliefs differ in degree of power and intensity; some beliefs are more central to a person's core than others; and the more central a belief is to one's core, the more difficult it is to change that belief. Beliefs that are central to one's core are those beliefs that define one's identity and beliefs that one shares with others. Rokeach also distinguished between derived beliefs, those that one learns from others, and underived beliefs, those that one learns through personal experience. Underived beliefs are more central to one's core than derived beliefs.

Nisbett and Ross (1980) also discuss the idea of a belief system, or beliefs, that are central to one's core. They coined the term "the perseverance phenomenon of theory maintenance" and hypothesized that the earlier a belief is acquired, the more difficult it is

to alter. Additionally, these early beliefs affect perception and continually influence the processing of new information. According to Nisbett and Ross (1980), the perseverance phenomenon often leads people to cling to old beliefs even when they are faced with evidence contrary to their beliefs.

Posner, Strike, Hewson and Gertzog (1983) used Piaget's concepts of assimilation and accommodation in their description of beliefs and how beliefs change. Assimilation is the process by which new beliefs are integrated into an existing system of beliefs. Accommodation is the process that occurs when new beliefs cannot be assimilated, thereby requiring that old beliefs be replaced or reorganized. Accommodation is a much more difficult process than assimilation, and, in order for it to occur, individuals must be displeased with existing beliefs and feel that the new beliefs are plausible and congruent with other factors being taken into consideration.

In "The role of beliefs in the practice of teaching," Jan Nesper (1987) outlined how beliefs are characterized into organized systems according to four features: existential presumption, alternativity, affective and evaluative loading and episodic structures. Existential presumptions emerge from the reification of abstract concepts into defined, stable, concrete entities. A classroom example of this occurs when a teacher labels his or her students, and these labels change from mere descriptive terms to actual entities thought to be embodied by the students. Often, these entities seem immutable and beyond the teacher's control. Alternativity occurs when one tries to create an idealized situation which may differ significantly from current realities. For instance, one tries to create a classroom environment that they have idealized in their minds but with which

they have had no direct experience. Nespor (1987) also suggested that beliefs are strongly affective and evaluative and that they often operate independently of other forms of cognition typically associated with knowledge. An example of the affective nature of beliefs occurs when teacher expectations are influenced by teachers' sometimes unrecognized feelings about students. An example of the evaluative nature of beliefs is illustrated by the fact that teachers' conceptions of subject matter knowledge is often influenced by the importance teachers place on the course content. Episodic structures are beliefs that are stored and organized in terms of personal experiences or episodes. An example of an episodic structure occurs when teachers are asked to explain why they became teachers. In response, many teachers describe a crucial experience or an influential teacher that produces a highly-detailed episodic memory that has served to guide them on this career path.

Goodman (1988) and Calderhead and Robson (1991) also noted the episodic nature of beliefs. Goodman (1988) found that teachers used "guiding images" from their past to create "intuitive screens" through which they sorted new information. Calderhead and Robson (1991) discovered that pre-service teachers use images of teaching from their past experiences as students to interpret happenings in their own classrooms and to determine courses of actions.

Nespor (1987) asserts two features of belief systems; they are non-consensual and unbounded. They are non-consensual because while knowledge is founded upon established canons of argument, beliefs are reached by each person individually through affective feelings, vivid personal memories and assumptions about entities and alternative



realities; none of which is open to outside evaluation or critical examination. Belief systems are also unbounded because there are no clear logical rules for ascertaining the relevance of beliefs by real-world standards, situations or events.

According to Nespor's (1987) study, belief systems are important to teachers because they are used by teachers to define tasks and to select strategies to use to accomplish tasks. Furthermore, the affective components of beliefs facilitates their storage in long-term memory because they take the form of gestalts that can be easily organized for representation and retrieval and acquire a "signature feeling." Nespor (1987) also argues that beliefs are very important to teachers because of the nature of their work. Teachers often have to deal with problems that are ill-defined and deeply entangled. Nespor (1987) described ill-defined problems as those problems where the goal is not clearly defined and/or the way to attain the goal(s) is not certain. Ill-defined problems are also those problems that require a person to go beyond the readily available knowledge sources and use background knowledge, assumptions and guesses to identify information relevant to the problem. Additionally, at various points of problem analysis, numerous alternative courses of action are available as possible solutions; the difficulty is in picking one alternative course of action. In these cases, research-based knowledge and academic theory are of limited use.

King and Kitchener (2002) examined ill-defined problems and developed The Reflective Judgment Model. The three period, seven stage model represents varying ways people justify beliefs. The first period, pre-reflective thinking, posits that knowledge is transferred from an authority figure or through first hand observation. In stage one,

individuals believe that knowledge is absolute, concrete and ascertained through observation. In stage two, knowledge is still absolute and certain, but not always readily available. In stage three, knowledge is still absolute, but sometimes temporarily uncertain. When knowledge is temporarily uncertain, personal beliefs are relevant until more absolute knowledge is available.

Period two is known as the quasi-reflective period, and individuals in this period begin to acknowledge the uncertainty of knowledge. Stage four of this period is characterized by the acknowledgement of the uncertainty of knowledge, but the uncertainty is attributed to misinformation or missing data. Individuals in this stage are still trying to force knowledge into pre-existing schema. In stage five, individuals begin to realize that knowledge varies depending on context and upon the point of view of the person sharing the information (King & Kitchener, 2002).

The last period, the reflective period, is denoted by the acceptance of the fact that knowledge claims cannot always be made with certainty. In light of this comes the ability to make decisions based upon the best evidence available. Individuals in this period are willing to reevaluate their own beliefs as new information becomes available. In stage six, an individual is able to construct a personal point of view on an ill-defined problem based on evidence from various sources while at the same time keeping in mind the personal perspectives of those sources. In stage seven, knowledge is believed to be a result of the inquiry process one undergoes when confronted with an ill-defined problem; knowledge is constructed from experiences, evidence and the perspectives of others (King & Kitchener, 2002).

The aforementioned researchers' work is about teacher beliefs in general. There is, however, a research agenda that addresses content-specific beliefs: the teaching of literature (Grossman, 1989), math (Freeman & Porter, 1989), science (Hollon, Anderson, & Roth, 1991) and history (Wilson & Wineburg, 1988). There is also a vast body of research on the beliefs of pre-service teachers. Studies in this area show that most individuals arrive to teacher education with well-formed beliefs about the nature of teaching, students and classrooms (Feiman-Nemser, McDiarmed, Melnick & Parker, 1988; Weinstein, 1989). These beliefs have been shaped by the thousands of hours pre-service teachers spent as students in their own elementary and secondary education (Tabachnick & Zeichener, 1984), and often, teacher education does little to alter these beliefs (Tabachnick & Zeichener, 1984; Zeichener, 1989). These pre-existing beliefs often act as filters, or intuitive screens (Goodman, 1988), through which pre-service teachers view and interpret the classroom teachers they observe during their practicum experiences (Calderhead, 1988).

Within the last two decades, much of the research on teacher beliefs has come from scholars in the field of multicultural education. Multicultural education theorists are interested in how the interaction of social class, gender, race, ethnicity, sexual identity and physical ability influences education (Banks, 2003b; Sleeter, 1991). The emphasis that theorists give to each of these factors varies, however. Critical pedagogy, for example, focuses upon issues of class, corporate power and globalization (Sleeter, 2001); critical race theory concentrates on race as well as classism, sexism and other forms of

oppression (Ladson-Billings, 1999); antiracist education challenges systematic racism in education (Dei, 1999).

Central to most conceptions of multicultural theory is an understanding of positionality (Nieto, 1992), the “concept that acknowledges that we are all raced, classed and gendered and that these identities are relational, complex, and fluid positions rather than essential qualities” (Martin & VanGuten, 2002, p. 46). Furthermore, multicultural theorists assert that one's identities, particularly those associated with one's positioning in society, influence the way in which one perceives and understands the world. In the classroom, positionality can influence how a teacher teaches and how a teacher perceives his or her students (Darling-Hammond, L., French, J. & Garcia-Lopez, S.P., 2002). One of the goals of multicultural theory is to facilitate the construction of knowledge from multiple positions to create culturally representative and equitable education for all (Martin & VanGuten, 2002).

### *2.2.1 Teacher Efficacy in Albert Bandura's Social Cognitive Theory*

Another construct of teacher beliefs that has generated much research is teacher efficacy which Tschannen-Moran and Woolfolk Hoy (2001) define as a teacher's “...judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who may be difficult or unmotivated” (p.3). Teacher efficacy is gaining renewed interest as an important psychological construct in understanding teacher motivation and teacher effectiveness. The construct has been discussed and measured for more than 20 years; however, there has been a resurgence of interest in its meaning and measure (Henson, 2001).

The teacher efficacy construct has a number of origins, but the majority of the research in this area has been based on the social cognitive theory of Albert Bandura (Henson, 2001). Bandura (2001) notes that “[a]mong the mechanisms of personal agency, none is more central or pervasive than people’s belief in their capacity to exercise some measure of control over environmental events...It is partly on the basis of efficacy beliefs that people choose what challenges to undertake, how much to expend in the endeavor, how long to persevere in the face of obstacles and failures, and whether failures are motivating or demoralizing” (p. 10).

In his social cognitive theory, Bandura argued against behaviorist theories of learning that focused on changing behavior through positive and negative reinforcers or consequences because this idea makes no reference to the conscious involvement of the individual. Bandura believed that cognitive processes play a part in the course of behavioral change in that an individual’s cognitive representations of anticipated outcomes can motivate behavior. These cognitive representations create the expectation in an individual that by behaving in a certain way, he or she can create positive outcomes or preclude problems. Reinforcements that result from carrying out the represented action are seen as motivating rather than simply strengthening an automatic behavioral response (Bandura, 1977).

In addition to these cognitive representations serving as motivators, Bandura also felt that goal setting and self-evaluation worked as cognitively based motivators. In this view, individuals are motivated by standards of performance that they set for themselves and are rewarded by attaining the self-determined level of performance. If an individual

realizes through self-evaluation that a discrepancy exists between what he or she hoped to accomplish and what was actually accomplished, the individual is motivated to take action to improve performance. Once the desired level of performance has been achieved, new goals are set for even higher levels of achievement. Bandura assumed that expectations of personal mastery had an effect on both the initiation and the persistence of attempts to reach a certain level of performance.

### *2.2.2 Sources of Teacher Efficacy*

According to Bandura's (1977, 1997) theory, expectation of personal mastery, or efficacy expectation, is based on four sources of information: enactive mastery experiences, learning through vicarious experiences, verbal persuasion from significant others and physiological and affective states. The most important of these is enactive mastery experiences, or performance accomplishments. As individuals experience success in particular situations, expectations are raised with regard to future success in similar situations. Enactive mastery experiences provide the individual with the evidence that he or she "can do whatever it takes to succeed" (1997, p. 80). Repeated failures, on the other hand, lower these efficacy expectations. Individuals create self-knowledge structures with regard to their efficacy beliefs. These beliefs are tested each time the individual experiences a situation relating to that efficacy belief. Task difficulty also plays a part in the development of efficacy beliefs. Sometimes, if an individual is faced with a very challenging task, they may be successful, but the effort expenditure and difficulty of undertaking the task may actually lower efficacy beliefs and leave the individual "...shaken rather than emboldened" (1997, p. 83).

The difficulty level of a task is also measured comparatively. If an individual is successful at a task while expending less effort than others completing the same task, self-efficacy is raised. Conversely, if the individual must expend far more energy to complete the task than others, self-efficacy beliefs are weakened. Additionally, self-efficacy can be raised if individuals attend to successes more than failure, and may be lowered if individuals attend more to poor performance (Bandura, 1997).

The second source of efficacy information is learning through vicarious experiences, or modeling. Seeing someone else perform a task gives an individual the feeling that they, too, can achieve at the same levels because individuals assess their capabilities in relation to what others have achieved. Models that make the most impact on efficacy beliefs are those that are judged to be similar in performance capabilities as well as in social and personal characteristics. For example, a woman of average physical fitness would increase her feelings of efficacy with regard to physical abilities if she witnessed another female with the same level of fitness achieve a goal more than she would if she chose as her model a professional female athlete or an average male athlete. Modeling is also more effective if an individual observes multiple models rather than a single one. Competent models also exert more influence than incompetent ones (Bandura, 1977, 1997).

Verbal persuasion from a credible other is also a source of efficacy information. The support of another person may boost an individual's sense of efficacy enough to provide motivation to persist in a difficult task and ultimately succeed at that task. Evaluative feedback that focuses on an individual's capabilities to achieve a task,

especially during skill development, raises efficacy beliefs. Feedback that focuses on the individual's effort expenditure may raise efficacy beliefs in the short term, but may lead the individual to believe that they succeeded merely through sheer effort rather than actual ability. This will lower efficacy beliefs in the long term. Verbal persuasion is most effective if it is given shortly after the task is successfully performed and if the task assigned is only moderately beyond the individual's current performance level. Bandura also noted that efficacy is elevated more when there is not a huge disparity between an individual's own beliefs about his or her capabilities and what they are told about their capabilities by the credible other (1997).

The final sources of efficacy information are an individual's physiological and affective states. In stressful situations, an individual's heart rate may increase, and they may perspire and have difficulty controlling their breathing. Depending upon the context in which these physiological activities occur, the individual may relate the causes to different sources. For instance, an inexperienced swimmer may attribute fatigue during an open water swim to lack of training or preparation, while a more experienced swimmer may attribute fatigue to strong currents and ocean swells. The impact of the physiological state will affect these two swimmers differently; the inexperienced swimmer will doubt his or her ability while the experienced swimmer will remain confident and blame environmental conditions. Affective features, such as mood, can also play a part in efficacy beliefs. Events and circumstances are processed differently when one is sleepy or sad rather than well rested and content (Bandura, 1997).



### *2.2.3 Methods of Measuring the Construct of Efficacy*

Attempts to measure teacher efficacy have come in many forms and have been grounded upon such theories as Rotter's (1966) social learning theory, Weiner's (1979, 1992) attribution theory and Bandura's (1977) social cognitive theory. This has been problematic as most measures of teacher efficacy reveal a two-factor structure when subjected to factor analysis. The meaning of these two factors does not seem to align consistently with one individual theory. In addition to these issues of validity, issues of reliability have also hindered measures of teacher efficacy. Another unresolved issue in measurement instruments of teacher efficacy is the fact that efficacy is context and subject matter specific. For instance, a language arts teacher may feel very efficacious in his or her ability to teach literature and grammar to struggling students but feel less efficacious when teaching reading, poetry, or writing to gifted students. While powerful in finding significant results, attempts to limit the scope of context in efficacy measures also reduce the generalizability and predictive value of the instrument.

The first attempt to measure a teacher's perceptions of his or her own capabilities came with a RAND commissioned study of teacher characteristics and student achievement (Armor et al., 1976). Researchers distributed a lengthy questionnaire to teachers, and two questions based upon Rotter's (1966) social learning theory emerged among the most powerful factors. The researchers created two questionnaire items grounded in Rotter's concept of locus of control. Teachers who indicated strong agreement with item one, "When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home

environment,” indicated a strong external locus of control believing that the influence of teachers has little influence in comparison to what happens outside of school. Since this instrument was developed, the concept of a teacher’s beliefs about the power of the influence of teachers and schools versus the power of the influence of variables outside of school such as socio-economic status, race, gender, home environment, has come to be known as general teaching efficacy (GTE) (Ashton, Olejnik, Crocker, & McAuliffe, 1982).

Teachers who agreed strongly with item two, “If I try really hard, I can get through to even the most difficult or unmotivated students,” indicated a strong internal locus of control believing that their individual ability as a teacher could overcome factors outside of school. This aspect of efficacy is referred to as personal teaching efficacy (PTE). The sum of the two items became an overall measure called teacher efficacy. The RAND study found a significant relationship between teacher efficacy and teachers’ success in teaching reading to minority students in urban schools (Armour et al., 1976).

The promising implications of the RAND study encouraged other researchers to continue investigating the concept of teacher efficacy, but they were concerned about the reliability of the measure as it comprised only two items. Still using Rotter’s idea of locus of control as their foundation, researchers tried to develop more comprehensive instruments to measure teacher efficacy. Two such instruments were Guskey’s (1981) responsibility for student achievement (RSA) and Rose and Medway’s (1981) teacher locus of control.

Guskey's (1981) RSA was a 30-item instrument which gave scenarios such as "When a student does well in your class, would it probably be because..." and the teacher filling out the form was required to distribute points between two alternatives. One alternative ascribed the occurrence to the actions of the teacher, such as "the encouragement you offered;" the other alternative indicated that the occurrence was related to the factors outside of the teachers' control, such as "that student had the natural ability to do well." Guskey framed the alternatives using Weiner's (1979) attributional theory and created responses that indicated that the occurrence happened due to specific teaching abilities, the effort put into teaching, the task difficulty, or luck. The instrument measured three factors: teacher responsibility for student success (R+), teacher responsibility for student failure (R-) and the combined score of the two which Guskey called responsibility for student achievement (RSA). Guskey found a strong correlation between scores on the teacher efficacy scale as developed by the RAND researchers and teacher responsibility for student success (R+) and student failure (R-). Guskey also found strong inter-correlations between overall RSA and responsibility for student success (R+) and student failure (R-).

The Rose and Medway (1981) teacher locus of control (TLC) was very similar to the Guskey instrument except that it was a 28-item instrument, and rather than measuring teacher responsibility for student success (R+) and student failure (R-), half of their items were created to describe student success and half to describe student failure. The two choices given to each item attributed the happening either to the teacher internally or to forces outside of the teacher. If the scenario was one of success and the teacher attributed

it to him or herself, it was rated as (I+); if the situation was a statement of student failure and the teacher attributed it to him or herself it was rated as (I-). The TLC has been significantly related to the RAND items, but the correlation was weak (Coladarci, 1992).

Neither the Guskey nor the Rose and Medway measures were used extensively by other researchers in published literature (Tschannen-Moran & Woolfolk Hoy, 2001). Other measures, based upon Bandura's (1977) social cognitive learning theory, were also being developed in the early 1980's. Ashton, Buhr and Crocker (1984) devised the Ashton vignettes that were short descriptions of situations teachers might face in their everyday work. Teachers were asked to assess how well they thought they would handle each situation and rate it on a Likert scale from "extremely effective" to "extremely ineffective." Then they were asked to rate how they would perform in each situation in comparison to other teachers. The comparison question scores were strongly correlated to the two RAND items, but there was no correlation between the RAND items and the self-rating items. The Ashton vignettes, much like the instruments grounded in Rotter's work, did not gain wide acceptance. One instrument that did gain acceptance, however, was the Gibson and Dembo Teacher Efficacy Scale (TES) (1984).

Gibson and Dembo's TES (1984) is a 30-item measurement upon which teachers rate statements such as "When a student gets a better grade than she usually gets, it is usually because I have found better ways of teaching" on a 6-point Likert scale from strongly disagree to strongly agree. This scale has been the most widely used measure of teacher efficacy in most literature to date despite the fact that researchers have raised

questions about the interpretation of the scale (Tschannen-Moran et al., 1998; Woolfolk & Hoy, 1990; Soodak & Podell, 1993).

#### *2.2.4 The Teacher Efficacy Scale and Problems of Interpretation*

Gibson and Dembo (1984) subjected the instrument to factor analysis and found that the items loaded on two factors. Because the instrument is grounded in Bandura's social cognitive learning theory, Gibson and Dembo assumed that the two factors were related to self-efficacy, or PTE, and outcome expectancy, or GTE. Other researchers have corroborated the two-factor structure of the instrument; however, upon closer analysis, several inconsistencies begin to emerge.

Soodak and Podell (1993) discovered that many of the items loaded on both factors. Because of this, they used a shortened 16-item measure and used only the items that loaded on one factor. When using the shortened version, Soodak and Podell (1993) found that items that originally loaded on the GTE factor now loaded on the PTE factor, and some items did not load strongly on either factor. The instability of the factor structure calls into question the interpretation of the meaning of the two factors. The second factor, GTE, had been linked to Bandura's outcome expectancy. Bandura (1986), however, states that outcome expectancy has little to do with what a person thinks others would be able to accomplish given a certain situation but rather what one believes he or she can expect based upon his or her own capabilities and performance. Therefore, Gibson and Dembo's labeling of GTE as an outcome expectancy does not truly align with Bandura's theory; it seems more closely related to Rotter's concept of external control.

Guskey and Passaro (1994) noted that all of the items of the TES relating to PTE were worded positively and reflected an internal orientation, while items related to GTE were worded negatively and reflected an external orientation. In an attempt to clarify the meaning of the two factors, they reworded the items so that for each category half of the statements were worded positively and half were worded negatively. Results of the analysis divided into internal/external categories rather than personal/general teacher dichotomies further complicated the interpretation of the two factors.

Despite the questionable interpretation of the two factors, Gibson and Dembo's TES (1984) is the most widely used measure of efficacy. It has also been recreated in various forms in an effort to address the fact that efficacy is context and subject matter specific. Attempts to limit the scope of measures of efficacy to specific subject areas and tasks have resulted in a few of the following instruments: the Science Teaching Efficacy Belief Instrument (Riggs & Enoch, 1990); an instrument measuring teacher efficacy for classroom management (Emmer, 1990); and the Dutch teacher efficacy scale for efficacy related to special education (Meijer & Foster, 1988). Bandura, too, attempted to create a more specific instrument that measured teacher efficacy across a wide variety of tasks: efficacy to influence decision making and school resources, efficacy for instruction, discipline, enlisting parent and community involvement and efficacy to create a positive school climate (1997). Unfortunately, information about the instrument's validity and reliability has not been made available.

In the 1990s, it seems that research on efficacy shifted. Researchers began pondering the implications of efficacy on reform efforts and questioning how the

construct of efficacy was measured. Individual teacher efficacy appeared less and less in the literature as many researchers turned their attention to collective teacher efficacy. As Tschannen-Moran et al. (1998) noted:

This appealing idea, that teachers' beliefs about their own capabilities as teachers somehow matter, enjoyed a celebrated childhood, producing compelling findings in almost every study, but it has also struggled through the difficult, if inevitable, identity crisis of adolescence...teacher efficacy now stands on the verge of maturity... (p. 202)

The promise of the concept of teacher efficacy as a potential tool of reform has served as an impetus for researchers to continue searching for an adequate instrument to measure efficacy.

#### *2.2.5 Summary of Teacher Beliefs*

In an effort to seek ways to improve adolescent literacy, educational researchers are now looking beyond student-level variables to school-level variables that may increase student achievement. One such variable is the classroom teacher, and a promising area of study with regard to the classroom teacher is teacher beliefs. Rokeach (1968) defined a belief system as “having represented within it, in some organized psychological but not necessarily logical form, each and every one of a person's countless beliefs about physical and social reality” (p.2). He also posited that beliefs differ in degree of power and intensity, that some beliefs are more central to a person's core than others and that the more central a belief is to one's core, the more difficult it is to change that belief. Nisbett and Ross (1980) and Posner, Strike, Hewson and Gertzog

(1983) offered further support for the idea that once beliefs are attained, it is very difficult to change them.

Jan Nespör (1987) outlined how beliefs are characterized into organized systems according to four features: existential presumption, alternativity, affective and evaluative loading and episodic structures. Goodman (1988) and Calderhead and Robson's (1991) work verified the existence of episodic structures as a way of characterizing beliefs. Nespör (1987) noted that belief systems are important to teachers because belief systems are used by teachers to define tasks and select strategies to use to accomplish tasks. Furthermore, teachers often have to deal with problems that are ill-defined and deeply entangled. In these instances, teachers must go beyond readily available knowledge sources and use background knowledge, assumptions and guesses to identify information relevant to the problem. King and Kitchenner (2002) examined ill-defined problems more closely and developed a three period, seven stage model, The Reflective Judgement Model, to represent varying ways people justify beliefs.

The aforementioned researchers' work is about teacher beliefs in general. There are, however, research agendas that address more specific issues regarding teacher beliefs: content-specific beliefs, the beliefs of pre-service teachers and how the interaction of social class, gender, race, ethnicity, sexual identity and physical ability influences teacher beliefs. Another specific area of teacher beliefs is research in the concept of teacher efficacy.

Teacher efficacy is a teacher's "...judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who



may be difficult or unmotivated” (Tschannen-Moran & Woolfolk Hoy, 2001, p.3). The concept is grounded in Albert Bandura’s (1977) social cognitive theory. According to Bandura’s (1977, 1997) theory, expectation of personal mastery, or efficacy expectation, is based on four sources of information: enactive mastery experiences, learning through vicarious experiences, verbal persuasion from significant others and physiological and affective states.

Attempts to measure teacher efficacy have come in many forms: a questionnaire developed by RAND researchers (Armor et al., 1976), Guskey’s (1981) responsibility for student achievement (RSA), Rose and Medway’s (1981) teacher locus of control, the Ashton vignettes (Ashton, Buhr & Crocker, 1984), Gibson and Dembo’s Teacher Efficacy Scale (1984), as well as some more specific instruments of measurement such as the Science Teaching Efficacy Belief Instrument (Riggs & Enoch, 1990), an instrument measuring teacher efficacy for classroom management (Emmer, 1990) and the Dutch teacher efficacy scale for efficacy related to special education (Meijer & Foster, 1988).

### 2.3 The Ohio State Teacher Efficacy Scale

In an attempt to create an instrument to measure teacher efficacy sensitive to the fact that teachers must complete a variety of tasks in circumstances that may either be supportive or obstructive, Tschannen-Moran and Woolfolk Hoy (2001) convened a seminar of eight graduate students, all of whom had teaching experience, at The Ohio State University. The seminar was entitled *Self-efficacy in Teaching and Learning*. The seminar students studied the various instruments used to measure teacher efficacy and concluded that they thought that the unpublished efficacy measure created by Bandura

(1997) was the best instrument upon which to model their own. Each seminar participant selected items from Bandura's instrument that he or she felt measured a task important in the daily life of a teacher. Additionally, each member created 8 to 10 more items that they felt were overlooked by Bandura's instrument. In the end, the seminar group compiled all of their ideas, pared down items where there was much overlap, discussed each item, revised where necessary and ended up with a 52-item instrument. Items were to be rated by teachers on a 9-point Likert scale.

The seminar members and two researchers conducted three studies with the new scale which they called the Ohio State Teacher Efficacy Scale (OSTES). Two-hundred forty-four participants took part in the first study. The group was comprised of 78 in-service teachers and 146 pre-service teachers who were taking classes at The Ohio State University. Participants not only responded to the 52 items on the OSTES, but they were also asked to rate the importance of each item for effective teaching by indicating 1- not at all, 2 – somewhat, 3 – important, or 4 – critical. The participants marked all items as being either important or critical, so none of the 52 items was dropped from the instrument.

The items were subjected to principal-axis factoring with varimax rotation, and 10 factors emerged. The first factor had an eigenvalue of 20.7 and accounted for 39.9% of the variance in respondents' scores. To select items for further analysis, the researchers chose items with loadings higher than 0.60; this yielded 31 items. An additional item was also selected even though it only had a factor loading of 0.59 because it addressed motivation and teaching, and the seminar group believed it was a critical task of teaching

that was not covered fully enough by the 31 other items. Hence, the first study narrowed the original 52-item instrument into a 32-item instrument.

The second study included 217 participants, 70 pre-service teachers, 147 in-service teachers with a mean of 8.5 years of teaching experience and 3 participants who failed to indicate their teaching status. The participants were students enrolled in coursework at three universities. Each participant filled out the 32-item OSTES.

Factor analysis of the 32-item instrument yielded eight factors with eigenvalues greater than one and accounted for 63% of the variance in the participants' scores. A scree test revealed that two or three factors could be extracted, so the seminar group and researchers discussed each solution. The two-factor solution revealed that items related to classroom management loaded across both factors and loadings were low. In the three-factor solution, classroom management emerged as a separate factor and the other two factors were more clearly defined. Because of this, the seminar group decided that the three-factor solution was parsimonious and interpretable.

Upon further analysis of the three-factor solution, the seminar group removed items with the lowest loadings, factors that loaded on more than one factor and factors that seemed superfluous. This reduced the instrument from 32 to 18 items. The final three factors were labeled efficacy for student engagement (8 items), efficacy for instructional strategies (7 items) and efficacy for classroom management (3 items) and accounted for 51% of the variance in respondents' scores. Subscale scores for each factor were computed by calculating the mean of the responses to the items retained within each

factor. Reliabilities for each factor were 0.82 for engagement, 0.81 for instruction and 0.72 for management.

The researchers conducted a second-order factor analysis using the responses from both study 1 and study 2. This analysis revealed one strong factor with high factor loadings ranging from 0.48 to 0.84. The reliability for the 18-item measure was 0.95. The researchers also found moderate positive correlations between the three subscales. These findings suggest that the 18-item instrument is a valid measure of the construct of efficacy.

To further test the validity of the 18-item instrument, correlations between the new instrument and existing instruments were determined. Participants in study 2 were also asked to respond to the RAND items and to the Hoy and Woolfolk (1993) adaptation of the Gibson and Dembo TES. Scores on the OSTES were positively correlated with both measures. In addition to responding to other efficacy measures, the participants were also asked to fill out the Willower, Eidell and Hoy's (1967) pupil control ideology form and Forsyth and Hoy's (1978) work alienation scale. These scales were used because both work alienation and pupil control ideology that indicates a custodial rather than humanistic perspective of students are counter indicative of high teacher efficacy. Correlation tests found that teacher efficacy was negatively related to both measures. The correlations were run again using only data from in-service teachers, and the results were similar.

The researchers were still concerned about the instrument because of the weakness of the classroom management factor. In a confirmatory analysis with 183 in-

service teachers, Roberts and Henson (2001) found the classroom management factor to be weak and recommended eliminating it from the instrument. The work of the seminar group, however, revealed that classroom management is considered essential to good teaching, so ways of strengthening the instrument were investigated. Because the instrument only contained three questions related to classroom management, it was determined that more items needed to be created to capture the construct of classroom management. The seminar group looked to Emmer's (1990) teacher efficacy for classroom management scale for ideas for more items. The seminar group also decided to include some items that had been overlooked by other measures of efficacy. For instance, most measures address the difficulties of working with hard to motivate students; none of the instruments address the challenges of working with capable students nor do they look at using instructional strategies to promote student thinking.

The instrument that resulted from this brainstorming session was field tested in a class of 19 people comprised of 17 teachers and two teacher educators. Feedback from this group helped to shape the revised instrument of 36 items. This instrument was tested in a study with 410 participants. Participants were 103 pre-service teachers, 255 in-service teachers with a mean score of 8.2 years of teaching experience and 38 participants who failed to indicate their status.

Four factors emerged from the principal-axis factor analysis with varimax rotation of the 36-item instrument which accounted for 58% of the variance in respondents' scores. A scree test revealed that three factors could be extracted, the same three factors that emerged from study two: efficacy for student engagement (12 items), efficacy for

instructional strategies (15 items) and efficacy for classroom management (9 items). The instrument items were reduced to 24 by selecting the eight items for each factor with the highest loadings. The 24 items were subjected to another factor analysis and the same three factors emerged. Efficacy subscale scores were also calculated for each factor, and reliabilities for the subscales were 0.87 for engagement, 0.91 for instructional strategies and 0.90 for classroom management. Intercorrelations for the subscales were calculated as well and were 0.58, 0.70 and 0.69, respectively ( $p < 0.001$ ).

Because of the high reliabilities of the subscales, researchers attempted to create an even more parsimonious instrument by reducing the number of items for each subscale from the eight items with the highest loadings to the four items with the highest loadings. It was determined that the factor structure remained unchanged and the reliabilities remained high with 0.81 for engagement, 0.86 for instructional strategies and 0.86 for classroom management. Hence, further analysis was conducted on both the short form and the long form.

Both forms were subjected to two factor analyses. One factor analysis was conducted using the responses of 111 pre-service teachers, and another factor analysis was conducted with the responses of 255 in-service teachers. In the study of the in-service teachers, the same three factors emerged for both forms and accounted for 54% of the variance in respondents' scores. Among pre-service teachers, however, the factor structure was not as distinct, so another principal-axis factoring was conducted calling for only one factor to be extracted. The analysis of the one factor accounted for 57% of the variance in respondents' scores on the 24-item instrument and 61% of the variance in

respondents' scores on the 12-item instrument. The authors of the instrument indicate that subscale scores may not be telling for pre-service teachers and that the total score may be a more appropriate measure of their efficacy (Tschannen-Moran & Woolfolk Hoy, 2001).

The researchers then combined the data from the pre-service and in-service teachers to conduct a principal-axis factoring of the three subscales. The 24-item instrument revealed one strong factor that accounted for 75% of the variance in respondents' scores. The 12-item instrument also revealed one strong factor and accounted for 68% of the variance in respondents' scores. This second-order factor and the moderate positive correlations of the three subscales indicated that the instruments do truly measure the construct of teacher efficacy and that the instrument could be used to calculate a total score as well as three subscale scores. The researchers ran another principal factor analysis specifying one factor to check the appropriateness of calculating one overall score for the instrument. They found that all items loaded on the one factor for both the short form and the long form. Reliability for the short form was 0.90, and reliability for the long form was 0.94.

The construct validity of both forms was also measured by determining the correlation of each form to other measures of efficacy. Participants in study 3 were also asked to respond to the RAND items and to the Hoy and Woolfolk (1993) adaptation of the Gibson and Dembo TES. Scores on both the short form and the long form of the OSTES were positively correlated with both measures.

The results of these studies indicate that the OSTES is a valid and reliable measure of teacher efficacy. Positive correlations between the OSTES and other

measures of efficacy provide further evidence of its construct validity. The OSTES, however, has the added benefit of being more specific to the tasks of teaching and to measuring the essence of what makes a teacher effective. The complete survey of the OSTES, now called the Teachers' Sense of Efficacy Scale, can be found in Appendix E.

### *2.3.1 Summary of The Ohio State Teacher Efficacy Scale*

There have been issues with reliability and validity of many measures of teacher efficacy, and most of the measures do not take into account the fact that efficacy is context and subject matter specific. In an attempt to overcome the problematic features of existing measures of teacher efficacy, Tschannen-Moran and Woolfolk Hoy created a The Ohio State Teacher Efficacy Scale (2001), a survey containing 24 items in a nine-point Likert format anchored from a score of 1 indicating "nothing" to a score of 9 indicating "a great deal." In addition to a total score for efficacy, the measure can be broken down into three subscales: efficacy for instructional strategies, classroom management and student engagement. Studies indicate that the OSTES is a valid and reliable measure of teacher efficacy, and the OSTES has the added benefit of being more specific to the tasks of teaching and to measuring the essence of what makes a teacher effective.

### 2.4 Teacher Efficacy and its Relationship to Various Aspects of Schooling

Teacher efficacy has been related to student outcomes such as motivation (Midgley, Feldlaufer & Eccles, 1989), achievement (Ashton & Webb, 1986) and students' own sense of efficacy (Anderson, Green & Loewen, 1988). Teacher efficacy has also been related to teacher behavior in the classroom. Allinder (1994) found that



teachers with high efficacy plan more thoroughly and are more organized. Highly efficacious teachers have a willingness to try new strategies (Guskey, 1988), persist when teaching becomes difficult (Coladaraci, 1992) and work with struggling students longer (Gibson & Dembo, 1984). Teachers with high efficacy show greater commitment to teaching (Evans & Tribble, 1986), more enthusiasm for teaching (Allinder, 1994) and are more likely to stay in the profession of teaching (Burley, Hall, Villeme & Brockheimer, 1991).

The relationship between teacher efficacy and student achievement is often indirect as highly efficacious teachers behave in certain ways, and those behaviors result in increased student achievement. Ashton and Webb (1986) found that teachers with a high sense of efficacy have high expectations for all students, establish classroom environments that encourage warm interpersonal relationships and promote strong academic work. They are more humanistic in their classroom management style (Woolfolk & Hoy, 1990), evoke trust from their students (DaCosta & Riordan, 1996) and favor student-centered classrooms (Czerniak & Schriver, 1994) as well as activity-based and experiential learning (Enochs, Scharman & Riggs, 1995). Furthermore, teachers with high efficacy are more likely to seek assistance from other educational professionals (DeMesuita & Drake, 1994) and promote parental involvement in schooling (Hoover, Dempsey, Bassler & Brissie, 1992).

#### *2.4.1 Teacher Efficacy and Student Achievement in Reading*

While there has been much examination of teacher efficacy and its relationship to various aspects of schooling, research that attempts to uncover to what extent teacher

efficacy predicts student achievement in reading is sparse. The first study to examine the impact of teacher efficacy on student achievement in reading was conducted in 1976 (Armor et al.). The purpose of the study was to identify school and classroom policies and other factors that might increase reading scores in inner-city children. Researchers in this study developed what has come to be known as the RAND measure of teacher efficacy, a two-item measure. The first item related to what now has come to be known as general teaching efficacy, or a teacher's belief about the power of external factors compared to the influence of teachers and schools in general on student learning. The second item related to what is now called personal teaching efficacy, or an individual teacher's perceptions of their own teaching abilities to influence student learning. In this study, the two responses were combined for an overall measure of teacher efficacy. Measures of student achievement in reading in 32 district junior high schools for students at both the sixth and seventh grade were obtained from scores on the California Test of Basic Skills (CTBS). Researchers reported finding a strong correlation between teacher efficacy and teachers' success in teaching reading to minority students in an urban context, although discussion of the analysis used to determine the correlation is not in the eighty-five page document.

Ashton and Webb (1986) studied 48 basic skills teachers with students in grades 9 - 11. Student achievement scores in mathematics, language and reading subtests of the Metropolitan Achievement Test (MAT) were obtained for the 1980 and 1981 school years for students who were struggling to pass and had been placed in special basic skills class for remediation and preparation for the test. Teacher efficacy was measured using a

variety of instruments: a questionnaire that contained the two RAND items, an eight-item forced-choice measure of efficacy that has come to be known as the Webb efficacy measure and a measure of personal teaching efficacy comprised of 15 vignettes now known as the Efficacy Vignettes. To examine the unique contribution to students' MAT reading achievement scores by teacher efficacy as measured by the RAND and the Webb efficacy instruments, a stepwise multiple regression was computed controlling for the 1980 school year scores. None of the correlations was found to be significant.

Tracz and Gibson (1986) also conducted a study into the effect of teacher efficacy on student achievement in reading. Their study measured efficacy with Gibson and Dembo's TES. Measures of student achievement were obtained from class mean standard scores on CTBS reading, math and language. The population for the study comprised 14 elementary school teachers and their students who were in grades 4 - 6. In this study, GTE and PTE were kept separate. In the analysis, these two factors correlated very little with each other ( $r = .22, p = .22$ ) and correlated very differently across the three achievement areas. The only strong correlation involving student achievement in reading was with PTE ( $r = .55, p = .02$ ). In stepwise multiple regression analysis with achievement scores as dependent measures, the variable of PTE accounted for a significant amount of variance in reading achievement ( $F = 5.13; df = 1, 12; p = .04; R^2 = .30$ ).

Anderson, Green and Loewen (1988) used the Gibson and Dembo instrument to find that PTE at the beginning of the school year predicted student achievement in language, reading and math in grade 3 but not in grade 6. In the analysis, the teacher was

assigned the class mean score for achievement. With the class as the unit of analysis, sample sizes were small, and the likelihood of finding statistical significance was greatly reduced. When a correlation was found between PTE and the achievement scores of third graders, two sets of stepwise multiple regressions were performed to determine which of the variables best accounted for differences in student achievement. Teachers' sense of personal efficacy accounted for 54% of the variance for grade 3 students ( $F = 9.39$ ;  $df = 1, 6$ ;  $p = .022$ ;  $R^2 = .84$ ).

Borton (1991) expanded upon the research of Gibson and Tracz (1986) in a study of 79 third and fourth grade students and their three teachers. Borton utilized the same measurement instruments, the Gibson and Dembo TES for teacher efficacy and CTBS scores for student achievement in reading. In a stepwise regression, neither teacher efficacy nor personal teaching efficacy emerged as significant predictor variables for student achievement scores on the reading portion of the CTBS. Hence, Borton's (1991) findings contradict those of Gibson and Tracz (1986).

A summary of the research on efficacy and reading achievement is provided in Table 1.

Table 1

*Review of Research on Teacher Efficacy and Student Achievement in Reading*

Study	Sample	Instrument(s)	Statistical Procedure	Findings
Armour et al., (1976)	Junior high teachers of minority students in urban context	RAND questionnaire CTBS	Correlation with multiple stepwise regressions	Significant correlation between teacher efficacy and reading achievement
Ashton & Webb (1986)	48 basic skills teachers of grades 9-11	Questionnaire with RAND items Webb Efficacy Scale Ashton Vignettes  MAT	Correlation with multiple stepwise regressions	No significant correlation between teacher efficacy and achievement
Tracz & Gibson (1986)	14 elementary teachers of grade 4-6	TES (Gibson & Dembo, 1984) CTBS	Correlation with multiple stepwise regressions	Significant correlation between personal teacher efficacy and achievement
Anderson, Green & Loewen (1988)	Elementary teachers of grades 3 and 6	TES (Gibson & Dembo, 1984) Canadian Achievement Test	Correlation with multiple stepwise regressions	Significant correlation between personal teacher efficacy and reading in grade 3 but not grade 6
Borton (1991)	3 elementary teachers of grades 4-6	TES (Gibson & Dembo, 1984) CTBS	Correlation with multiple stepwise regressions	No significant correlation between personal teacher efficacy and achievement

#### *2.4.2 Summary of Teacher Efficacy and its Relationship to Various Aspects of Schooling*

While there has been much examination of teacher efficacy and its relationship to various aspects of schooling, research that attempts to uncover to what extent teacher efficacy predicts student achievement in reading is sparse. To date, there have only been five such studies, and the findings of these five studies are inconclusive. While the Armor et al. (1979), the Anderson, Green and Loewen (1988) and the Gibson and Tracz (1986) studies found a correlation between teacher efficacy and student achievement in reading, the other two studies did not (Ashton & Webb, 1986; Borton, 1991). It is interesting also that the Anderson, Green and Loewen (1988) study found a correlation between PTE and reading achievement for third graders but not for sixth graders. Research has indicated that children experience a decline in interest and competence beliefs regarding reading during the transition to junior high school (Wigfield, Eccles, MacIver, Reuman, & Midgley, 1991; Oldfather & McLaughlin, 1993; Oldfather & Dahl, 1994). With the exception of the Ashton and Webb study (1986) which included basic skills students in grades 9-11, all of the studies focused on students in grade 7 or below. More investigation is needed into the relationship between teacher efficacy and the reading achievement of adolescent students. The five studies to date also utilized the two RAND items to measure efficacy, the Gibson and Dembo TES, or the Webb efficacy scale. Each of the instruments is founded upon the problematic concepts of personal and general teaching efficacy, rather than a more specific instrument that measures teacher efficacy across a wide variety of contexts and tasks.

## 2.5 The Three Sub-scales of The Ohio State Teacher Efficacy Scale

While the OSTES is not a measure of efficacy specifically designed for teachers of reading or the language arts, its three-component sub-scale structure lends itself nicely to the assessment of efficacy of language arts and reading teachers. In *Developing Engaged Readers in School and Home Communities* (1996), Guthrie, McGough, Bennett and Rice described engaged readers as motivated to read, strategic in methods of achieving comprehension, metacognitive in the construction of new understandings and socially interactive in their literacy endeavors. Similarly, Guthrie and Wigfield (2000) proposed that “engaged readers in the classroom or elsewhere coordinate their strategies and knowledge (cognition) within a community of literacy (social) in order to fulfill their personal goals, desires and intentions (motivation). The three-component sub-scale structure of the OSTES examines teacher efficacy in the same three areas: teacher efficacy for student engagement (motivation), instructional strategies (cognition) and classroom management (social). Hence, while the OSTES is not intentionally designed for English/Language Arts teaching, it does measure three areas specifically related to teachers’ capacity to create engaged readers. This chapter will now examine the research related to the topics of the individual subscales of the OSTES.

### *2.5.1 Teacher Efficacy for Student Engagement*

The OSTES is comprised of eight questions that assess teacher efficacy for student engagement:

1. How much can you do to get students to believe they can do well in schoolwork?
2. How much can you do to help your students value learning?
3. How much can you do to motivate students who show low interest in learning?
4. How much can you assist families in helping their children do well in school?
5. How much can you do to improve the understanding of a student who is failing?
6. How much can you do to help your students think critically?
7. How much can you do to foster student creativity?
8. How much can you do to get through to the most difficult students?

A thorough review of the literature showed that there was no research available specifically on teacher efficacy for student engagement; there is, however, research addressing the influence of teachers on student motivation in general. Many studies have documented associations between aspects of positive student-teacher relationships and student motivation in school. Goodenow (1993b) found that middle school students who felt that their teachers were supportive and caring had a stronger sense of belonging to the school community and, therefore, expressed higher levels of motivation and achievement as measured by teacher assessment of student effort and self-rated measures of student expectancies. Roeser, Midgley and Urdan (1996) conducted a follow-up study to Goodenow's (1993b), again with middle school students, and found that students' perceived relationship with their teachers was the single best predictor of a students' sense of belonging to the school community which in turn predicted increased motivation and achievement.



Wentzel (1998b) conducted a study of sixth graders and found that teacher support was independently related to many motivational variables even after controlling for peer and parental support. Wentzel (2002) later examined teachers' influence on student adjustment in middle school. Student adjustment was defined by students' academic and social interest in the teacher's class as well as students' academic performance and classroom behavior. Teaching dimensions that were investigated were fairness, teacher motivation, rule setting, negative feedback and high expectations. Wentzel's (2002) primary question was to what degree do teachers differ along these dimensions and whether or not these dimensions relate to students' school adjustment as measured by student motivation, classroom performance and behavior.

Wentzel (2002) focused on student motivation because scholars are increasingly recognizing the fact that students' success at school requires that students have a willingness to meet academic challenges. This willingness, or students' motivation outcomes, was identified under four categories: prosocial goal pursuit, responsibility goal pursuit, interest in class and mastery orientation. Prosocial and responsibility goal pursuit were defined as what students see themselves trying to accomplish.

Wentzel (2002) found that teachers did vary significantly along the five teaching dimensions. Correlations among variables indicated that the four motivational outcomes were significantly and positively related to the four teaching dimensions of fairness, teacher motivation, rule setting and high expectations and significantly and negatively related to negative feedback. Hierarchical multiple regressions indicated that the teaching dimensions accounted for significant amounts of variance in student motivation as

measured by the four outcomes of prosocial goal pursuit, responsibility goal pursuit, interest in class and mastery orientation, even after controlling for demographic variables.

High expectations was a positive, independent predictor for each outcome ( $\beta = .26, .31, .34,$  and  $.23$  respectively,  $p < .001$ ). Teacher fairness ( $\beta = .13, p < .05$ ) and teacher motivation ( $\beta = .27, p < .001$ ) were also positive predictors for interest in class. Teacher fairness ( $\beta = .16, p < .01$ ) was a positive predictor for mastery orientation. Negative feedback was a negative predictor for each of the four outcomes but was only of significance with regard to responsibility of goal pursuit ( $\beta = -.16, p < .01$ ). Wentzel's study (2002) confirms past studies that indicate that middle school students are cognizant of teacher behaviors that communicate caring and personal support (Wentzel, 1997), and, more importantly, that students' positive views of such teachers are related to increased interest and engagement in classroom activities (Wentzel, 1997, 1998b).

Murdock and Miller (2003) found the same to be true in their study of 206 eighth-grade students. Indicators of student motivation were measures of students' self-efficacy, students' intrinsic valuing of schooling and teacher-rated effort. Unlike the Wentzel (2002) study, Murdock and Miller (2003) controlled for both prior motivation and for parent and peer support. First, zero-order correlations demonstrated a relationship between perceived teacher caring and academic self-efficacy ( $r = .494, p < .01$ ), intrinsic valuing of education ( $r = .582, p < .01$ ) and teacher-rated effort ( $r = .345, p < .01$ ). To control for prior motivation and parent and peer support, Murdock and Miller (2003) conducted three hierarchical multiple regressions using students' eighth grade measures of self-efficacy, students' intrinsic valuing of schooling and teacher-rated effort as criterion

variables. In measures of self-efficacy, seventh-grade efficacy accounted for 10% of the variance, parent and peer support increased the accounted for variance by 25% and teacher caring accounted for an additional 4% of the variance. In measures of intrinsic valuing of schooling, teacher caring played a much more significant role. Prior motivation accounted for 9% of the variance and parent and peer support accounted for an additional 16% of the variance. Teacher caring accounted for an additional 14% of the variance. With regard to teacher-rated effort, prior motivation accounted for most of the variance, 15%. Parent and peer support accounted for an additional 8%, and teacher caring accounted for only a 2% increase in variance.

Therefore, it seems that student perception of teacher caring is influential with regard to students' intrinsic valuing of school. Teacher caring, in this study, was measured by students' reports of perceived expectations, perceived respect and perceived commitment to teaching. Evidence suggests that teachers do influence student motivation.

### *2.5.2 Teacher Efficacy for Instructional Strategies*

Eight questions on the OSTES assess teacher efficacy for instructional strategies:

1. To what extent can you use a variety of assessment strategies?
2. To what extent can you provide an alternative explanation or example when students are confused?
3. To what extent can you craft good questions for your students?
4. How well can you implement alternative strategies in your classroom?
5. How well can you respond to difficult questions from your students?
6. How much can you do to adjust your lessons to the proper level for individual students?
7. To what extent can you gauge student comprehension of what you have taught?
8. How well can you provide appropriate challenges for very capable students?

The OSTES survey spans a wide range of instructional strategies from methods of assessment and comprehension monitoring to explanation giving, questioning techniques and lesson planning. There is no research available on a teacher's ability to vary instructional strategies in general; most research, rather, investigates the effectiveness of the employment of a specific strategy in comparison to a control group where the specific strategy is not used. A review of all of the research of this type is beyond the scope of this review. There is, however, a body of research with regard to teacher efficacy and teacher willingness to learn and utilize new strategies and also about teacher efficacy and teacher willingness to differentiate instruction.

With regard to teacher willingness to learn and utilize new strategies, Guskey (1988) conducted a study to examine the relationship between teacher efficacy and teachers' attitudes toward the implementation of new instructional strategies. The sample included 120 elementary and secondary teachers who had just participated in a staff development workshop on mastery learning instructional strategies. Teacher efficacy was measured using the Responsibility for Student Achievement (RSA) scale (Guskey, 1981) as well as two efficacy items from the RAND Corporation's Change Agent Study (Berman & McLaughlin, 1977). Another section of the questionnaire was designed to assess teachers' attitudes toward the implementation of mastery learning instructional practices. Correlations revealed that more efficacious teachers (as measured by the RAND items) did tend to rate mastery learning as more important ( $r = 0.42$ ), more congruent with their present teaching practices ( $r = 0.36$ ) and less difficult to implement ( $r = -0.33$ ), than did their less efficacious colleagues. Therefore, Guskey (1988)

concluded that teachers with high levels of efficacy appear more likely to be receptive to the implementation of new instructional practices.

Albion (1999) investigated the link between teacher efficacy and the use of computers in teaching after reviewing several articles that identified a lack of confidence for teaching with computers as one of the main factors influencing the level of computer use by teachers with their students. To date, he could only find one study (Borchers, 1992) that tested the relationship. Borchers' (1992) study demonstrated that when teachers' self-efficacy belief in their ability to use computers was increased through staff development, they were more likely to integrate computers into their teaching strategies.

Much of the work on teacher efficacy and instructional strategies has been conducted in the field of special education. Wertheim and Leyser (2002) attribute this to their belief that the success of any special education legislation depends on the willingness and skill of teachers to make accommodations for individual student needs. Soodak and Podell (1993) were interested in examining teachers' willingness to work with exceptional education students. They created three case studies describing a second-grade male student who had behavioral and/or learning problems and asked 96 regular educators and 96 special educators to read the cases and determine whether the student was appropriately placed in a regular education classroom and whether they would refer this student to special education. Participants also completed Gibson and Dembo's (1984) TES questionnaire.

Soodak and Podell (1993) found that both special and regular educators were likely to agree with a regular class placement when they scored high on both dimensions

of teacher efficacy, personal and teaching. Regular educators with a higher sense of personal efficacy were more likely to agree with class placement than were regular educators with a lower sense of personal efficacy. Soodak and Podell (1993) hypothesized that regular educators who do not perceive themselves as being able to influence student outcomes believe that students with special problems should not be placed in regular classrooms. Allinder (1994) compared 73 special education teachers providing direct instructional services and 43 educators providing mostly indirect services to exceptional students and found that both teaching efficacy and personal efficacy were related to instructionally relevant effective teaching components.

Wertheim and Leyser (2002) conducted a study involving 191 pre-service teachers in Israel to determine whether their efficacy beliefs were related to their choices of instructional strategies as well as their perception of the effectiveness of the instructional strategies with regard to an inclusive environment. Fifty-three members of the sample were studying to teach at the early childhood education level, 57 at the elementary level and 81 at the junior high level. Each participant completed Gibson and Dembo's (1984) TES as well as a questionnaire about instructional strategies designed to measure pre-service teachers' perception of the acceptability of each strategy as well as their intent to use each strategy. Wertheim and Leyser (2002) obtained low but significant positive correlations between interns with high PTE and the intent to use individualized ( $r=.39, p=.001$ ) and diagnostic teaching strategies ( $r=.31, p=.001$ ), implement a variety of classroom management techniques ( $r=.28, p=.001$ ) and communicate with parents ( $r=.24, p=.001$ ), professionals ( $r=.22, p=.001$ ), students ( $r=.24, p=.001$ ) and principals

( $r=.18$ ,  $p=.001$ ) than did participants with lower PTE scores. The same pre-service teachers perceived these strategies to be more effective. Wertheim and Leyser's (2002) results corroborated the finding of Minke, Bear, Deemer and Griffen (1996) who found that teachers with high self-efficacy were more likely to concentrate on individualized instruction and to adapt teaching practices more readily.

It appears that teachers with high efficacy are willing to persist longer with students in regular education classrooms as well. Gibson and Dembo (1984) investigated if teachers of varying levels of efficacy exhibited different patterns of behavior with regard to academic focus, feedback and persistence in failure situations.

Methodologically, this was a case study of eight elementary school teachers, four with high efficacy and four with low efficacy. Gibson and Dembo used their own efficacy scale (1984) to measure efficacy. To measure academic focus, they utilized a teacher-use-of-time measure that yielded time allocation to daily rituals, transitions, whole class instruction, small group instruction, checking seatwork, preparation of paperwork, games, unfocused small talk and recess. To measure feedback, they used a question-and-answer feedback form that indicated when the teacher offered praise or criticism and when the teacher persisted even after a student failed to give a correct response by repeating the question, giving a clue or asking a new question. Not persisting was defined as giving the answer, asking another student and/or allowing another student to call out the answer.

Results indicated that high-efficacy teachers allocated 28% of their time to small group instruction, while low-efficacy teachers allocated 48% of their instructional time to small groups. It was observed, however, that low-efficacy teachers adhered to a much

more rigid small group format and appeared flustered if the format was not followed. High-efficacy teachers utilized the small group format with more ease and flexibility. It was also noted that more students remained on task in the groups in high-efficacy teacher classrooms than in low-efficacy teacher classrooms. High-efficacy teachers also spent considerable more time in whole class instruction. With regard to feedback patterns, high-efficacy teachers offered more praise for correct answers than did low-efficacy teachers, less criticism for incorrect answers and persisted longer with students who gave incorrect answers.

Saklofske, Michayluk and Randhawa (1988) conducted a study involving pre-service teachers who were engaged in a practicum experience. The purpose of the study was to determine if there was a relationship between efficacy and various aspects of the teaching evaluation rubric used to rate the interns at their practicum sites. Interns completed Gibson and Dembo's (1984) TES and results indicated that there was a small but significant positive correlation between PTE and lesson planning behaviors ( $r=.26$ ), classroom management behaviors ( $r=.23$ ) and questioning behaviors ( $r=.22$ ).

Therefore, it seems that there is evidence to suggest that teachers with high efficacy are more willing to implement new instructional strategies and technologies. They are also more willing to differentiate instruction and modify instruction for individual student needs whether the student is in exceptional or regular education. Furthermore, teachers with high efficacy appear to persist longer with students who give



incorrect responses and to praise students more often and criticize less. The impact of these findings on student achievement is unclear, however, and further research is needed.

### *2.5.3 Teacher Efficacy for Classroom Management*

The OSTES is comprised of eight questions that assess teacher efficacy for classroom management:

1. How much can you do to control disruptive behavior in your classroom?
2. How much can you do to get children to follow classroom rules?
3. How much can you do to calm a student who is disruptive or noisy?
4. How well can you establish a classroom management system with each group of students?
5. How well can you keep a few problem students from ruining an entire lesson?
6. How well can you respond to defiant students?
7. To what extent can you make your expectations clear about student behavior?
8. How well can you establish routines to keep activities running smoothly?

There is a small body of research on teacher efficacy and classroom management. The first research to appear in the literature was that of Barfield and Burlingame (1974) who found that teachers with a low sense of efficacy were custodial in pupil control ideology. Ashton and Webb (1986) noted that teachers with low efficacy were not only oriented toward control in their classrooms, they also tended to achieve control through punitive management strategies in comparison to high-efficacy teachers who encouraged student trust, autonomy and responsibility. Woolfolk and Hoy (1990) elaborated on these two studies and examined the relationship between prospective teacher's efficacy and their orientation toward pupil control and motivation. Teacher efficacy was measured using Gibson and Dembo's TES (1984). Pupil control ideology (PCI) was assessed with the PCI form developed by Willower, Eidell, and Hoy (1967) that conceptualizes PCI

along a continuum from custodial to humanistic. Custodial views of PCI are illustrated by a strong need to maintain order in a highly controlled setting. In its most extreme forms, the classroom atmosphere is one of “watchful mistrust” (Woolfolk & Hoy, 1990, p. 84). A humanistic perspective of PCI encourages students to learn through interaction and experience, and strict control is replaced by opportunities for students to demonstrate self-discipline. Motivational orientation of teachers was determined using the Problems in School Inventory, an instrument that contains eight vignettes, each with four possible solutions that range from highly controlling to highly autonomous (Deci, Schwartz, Sheinman, & Ryan, 1981).

Zero-order correlations revealed that teaching efficacy (TE) was significantly correlated with PCI ( $r = -.50, p < .01$ ), indicating that teachers with high TE favored a humanistic approach to PCI. Personal teaching efficacy (PTE) was not significantly correlated with PCI. Motivational orientation was not significantly related to TE or PTE. Through the employment of canonical correlations, it was found that TE and the TE x PTE interaction best predicted PCI. Regression analysis revealed a significant main effect for teaching efficacy and a significant interaction between TE and PTE on PCI. Teachers that rated high on both TE and PTE were more humanistic in their PCI than teachers with low PTE. However, when teachers are low on TE and high on PTE, they exhibit a more custodial approach to PCI.

The relationship of teacher efficacy and classroom management was examined more closely when Emmer and Hickman (1991) developed a separate self-efficacy scale to measure teacher efficacy for classroom management. The development of this scale

was prompted by the fact that a considerable amount of teachers' time is spent on behavioral outcomes that are not directly linked to student learning but rather to creating an environment in which student learning can occur. Only two items on the Gibson and Dembo (1984) scale relate to classroom management. Emmer and Hickman's (1991) scale for teacher efficacy in classroom management and discipline consists of three factors. Two factors borrow heavily from the Gibson and Dembo TES. The External Influences factor consists of variables outside of the teacher's control and resembles what Gibson and Dembo labeled TE, and Personal Teaching Efficacy on the Emmer and Hickman (1991) scale is created entirely with items originally on the Gibson and Dembo scale with the exception of one item. Emmer and Hickman's (1991) third factor is referred to as the Classroom Management/Discipline factor and consists of 11 items scored on a Likert scale that relate to teachers' self-ratings of efficacy in the area of classroom management such as "I have very effective classroom management skills" and "If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him quickly." Factor analysis supports that the Classroom Management/Discipline factor is distinct from other types of efficacy, and Emmer and Hickman (1991) report acceptable reliability for the factor (coefficient alpha .79 and test – re-test reliability .75).

Using the newly developed scale, Emmer and Hickman (1991) then sought to determine if teacher efficacy predicts preference for teaching strategies by presenting research participants with six vignettes detailing various student behavioral and academic problems. Each vignette was accompanied by 14 strategies, and subjects were asked to

indicate their likelihood of using each strategy on a 5-point response scale. The 14 strategies fell under three different response categories: Reductive Strategies indicated teachers' preference for using techniques such as time out, desists, consequences and warnings; Positive Strategies indicated teachers' preference for techniques that would establish or increase desirable behaviors by offering more praise, encouraging more effort, modifying assignments or teaching strategies and giving extra attention; and External Support which indicated teachers' preference for techniques that enlisted help from others in the form of obtaining more information about the student, referring the student to the office or enlisting the support of the student's peers. The Classroom Management/Discipline and Personal Teaching efficacy factors were positively correlated with Positive Strategies ( $r = .30, p < .05$  and  $r = .32, p < .05$  respectively), and the External Influences subscale was negatively correlated with preference for Positive Strategies ( $r = -.20, p < .05$ ). Emmer and Hickman's (1991) results corroborate the findings of Ashton and Webb (1986). Hence, it appears that there is a relationship between teacher efficacy and classroom management with regard to control orientation. Bandura (1993) noted that "teachers' beliefs in their personal efficacy to motivate and promote learning affect the types of learning environments they create and the level of progress their students achieve" (p. 117).

Bandura (1997), along with Lent, Brown and Hackett (1994) proposed that teacher efficacy plays a significant part in the implementation of academic choice. Several studies (Ryan, Connell & Grolnick, 1992; Weinert & Helmke, 1995; Miserandino, 1996) indicate that controlling environments, such as those sometimes

avored by teachers with a low sense of efficacy, reduce students' intrinsic motivation and sense of autonomy which contribute to low achievement, poor attitude toward school and anxiety (Boggiano & Katz, 1991). This is especially troublesome in light of research that suggests that students' perceived autonomy is related to a number of positive outcomes: enjoyment of school (Ryan & Deci, 2000), increased intrinsic motivation (Skinner, Wellborn & Connell, 1990) and perceived competence (Cordova & Lepper, 1996).

A recent study of teacher efficacy for classroom management (Morris-Rothschild & Brassard, 2006) investigated the relationship between efficacy for classroom management and teacher conflict management styles. Researchers measured efficacy for classroom management using Emmer and Hickman's (1991) scale and conflict management styles with the Rahim organizational conflict inventory-II (ROCI-II; Rahim, 1990). The ROCI-II delineates five distinct conflict management styles: an integrating style endorses a "win-win" attitude; a comprising style calls for sacrifice from both parties; an obliging style reflects a "lose-win" attitude indicating that the individual is more concerned about the other party's desires; a dominating style indicates that the individual's concerns override those of the other party's; and an avoiding style represents little concern for either party's goals or a strong desire to evade conflict altogether.

Morris-Rothschild and Brassard, (2006) discovered that teachers who were highly efficacious in classroom management favored the mutually focused classroom management styles of integration and compromise, indicating that they desired a proactive approach to handling student-teacher conflict. This finding is supported by past

research that reports that teachers with high efficacy are more willing to take personal responsibility for addressing students' behavioral issues with individual students than are teachers with low efficacy (Hughes, Grossman, & Barker, 1990). They are also less preoccupied with their own shortcomings, enabling them to engage in conflict management strategies that will benefit both themselves and their students (Bandura, 1980). High efficacy has also been associated with the ability to develop positive interactions with and among students (Rich, Lev & Fisher, 1996).

#### *2.5.4 Summary of the Three Sub-scales of The Ohio State Teacher Efficacy Scale*

Guthrie and Wigfield (2000) proposed that engaged readers coordinate their strategies and knowledge (cognition) within a community of literacy (social) in order to fulfill their personal goals, desires and intentions (motivation). The three-component, sub-scale structure of the Ohio State Teacher Efficacy Scale examines teacher efficacy in the same three areas: teacher efficacy for student engagement (motivation), instructional strategies (cognition) and classroom management (social). Hence, while the OSTES is not intentionally designed for English/Language Arts teaching, it does measure three areas specifically related to teachers' capacity to create engaged readers.

Research in the area of teacher efficacy for student engagement highlights the influence of teachers on student motivation. Many studies have documented associations between aspects of positive student-teacher relationships and student motivation in school. Students who feel they have a caring teacher have a stronger sense of belonging and have higher levels of motivation and achievement (Goodenow, 1993b). As a matter of fact, students' perceived relationship with their teacher is the single best predictor of

this phenomenon (Roeser, Midgley & Urdan, 1996). Wentzel's extensive work in the area (1997, 1998b, 2002) illustrates that teacher support is independently related to student motivation, classroom performance and behavior. Wentzel's research also confirms that middle school students are cognizant of teacher behaviors that communicate caring and personal support, and, more importantly, that students' positive views of such teachers are related to increased interest and engagement in classroom activities (Wentzel, 1997, 1998b). Murdock and Miller (2003) found the same to be true and also found that teacher caring plays a significant role in students' intrinsic valuing of schooling. The six studies reviewed in this section, however, comprised samples that included only middle school-aged students. More research needs to be done at the high school level. Additionally, contextual variables such as race, ethnicity and SES were not taken into consideration in any of the studies mentioned.

Research on teacher efficacy for instructional strategies presents evidence to suggest that teachers with high efficacy are more willing to implement new instructional strategies (Guskey, 1988) and technologies (Borchers, 1992). They are also more willing to differentiate instruction and modify instruction for individual student needs whether the student is in exceptional or regular education (Werthiem & Leyser, 2002; Soodak & Podell, 1993; Allinder, 1994). Furthermore, teachers with high efficacy appear to persist longer with students who give incorrect responses and to praise students more often and criticize less (Gibson & Dembo, 1984; Saklofske, Michayluk & Randhawa, 1988). The impact of these findings on student achievement overall and on achievement in reading specifically is unclear, however, and further research is needed.

Research on teacher efficacy for classroom management suggests that less efficacious teachers have a very custodial approach to classroom management (Barfield & Burlingame, 1974). Ashton and Webb (1984) and Woolfolk and Hoy (1990) corroborated this finding and added that such teachers also achieve control through punitive management strategies rather than through a more humanistic perspective. Environments created through such measures reduce students' intrinsic motivation and sense of autonomy which contribute to low achievement, poor attitude toward school and anxiety (Boggiano & Katz, 1991). Emmer and Hickman (1991) found that teachers high in personal teaching efficacy preferred positive strategies such as offering more praise, encouraging more effort, modifying assignments or teaching strategies and giving extra attention to students. In situations involving conflict in the classroom, highly efficacious teachers favored the mutually focused classroom management styles of integration, which endorses a "win-win" attitude, and compromise, which calls for a sacrifice from all parties (Morris-Rothschild, & Brassard, 2006). The impact of these positive strategies of classroom management on student achievement in reading, however, is yet to be determined.

## 2.6 Chapter Summary

This chapter reviewed important aspects of adolescent literacy: the development of literacy skills beyond emergent literacy and current trends in the study of adolescent literacy. The broad concept of teacher beliefs was then discussed prior to an in-depth explanation of Albert Bandura's social cognitive theory, a theory that has guided much of the research in one concept of teacher beliefs called teacher efficacy. Various methods of



measuring the construct of efficacy were reviewed as well as some of the problems associated with measuring teacher efficacy. This chapter then outlined the literature concerning the relationship between teacher efficacy and various aspects of school in general as well as the relationship between teacher efficacy and student achievement in reading specifically. Finally, since the OSTES, the instrument used to measure teacher efficacy in this study, divides the measure of efficacy into three sub-categories, teacher efficacy for instructional strategies, classroom management and student motivation, the literature for each sub-category was reviewed.

## CHAPTER 3

### METHODOLOGY

Descriptions of the research design, including the sample, data collection, dependent and independent variables and the instruments used to measure the variables, are provided in this chapter. Because the OSTES is a relatively new instrument, there is a thorough discussion of the development of the instrument as well as an outline for a confirmatory factor analysis that was conducted to verify construct validity. The contextual variables of prior achievement, SES, grade and ethnicity are also discussed and a justification for taking them into consideration is proposed. Finally, in the data analysis section, methodological points related to the use of hierarchical linear modeling, the primary method of analysis used in this study, are described and the models to be run are delineated. While interpretation techniques are included throughout the descriptions of the various groups of models to be run, more attention is given to interpretation techniques before the assumptions and limitations of the study are addressed.

#### 3.1 Purpose of the Study

The purpose of this study was to investigate the effects of teacher efficacy for instructional strategies, classroom management and student engagement on ninth and tenth grade students' reading achievement. Teachers with high efficacy were hypothesized to positively impact students' reading gains. Student contextual variables of

prior student achievement, socioeconomic status, ethnicity and grade were controlled for in the study.

### 3.2 Research Questions

The following research questions were addressed:

1. What is the influence of teacher efficacy regarding instructional strategies on student reading achievement gains?
2. What is the influence of teacher efficacy regarding classroom management on student reading achievement gains?
3. What is the influence of teacher efficacy regarding student motivation on student reading achievement gains?
4. What is the influence of total teacher efficacy (teacher efficacy regarding instructional strategies, classroom management and student motivation combined) on student reading achievement gains?

### 3.3 Research Design

#### *3.3.1 Sample*

The sample for this study was drawn from two large urban school districts on the west coast of Florida. Urban districts were selected to hold constant the differences that might occur between urban and non-urban districts. One school district included in the sample is the tenth largest school district in the United States, with a Pre-K to 12th grade student population of more than 180,000. The district includes twenty-five high schools, each with an average student population of 1,800 students per school. The second district is the twenty-third largest school district in the United States, with a Pre-K to 12<sup>th</sup> grade

student population of 152,000. This school district is not quite as large as the other with only 17 high schools, each with an average student population of 1,600 per school. Of the 312 teachers in the two districts who teach ninth or tenth grade English, 115 volunteered to fill out the Teacher Personal Information Questionnaire Information Questionnaire and the OSTES for a return rate of 37%.

*Teachers.* Of the 115 ninth and tenth grade English teachers who volunteered to fill out the researcher-created Personal Teacher Information Questionnaire and the OSTES, five teachers were retained for the confirmatory factor analysis but had to be eliminated prior to running the multivariate analysis for various reasons. Because the Office of Accountability could not provide the student data for four of the teachers prior to the running of the analysis, the data for those teachers were removed from the multivariate analysis. The fifth teacher was eliminated because no record of her could be found. It was determined that this teacher had likely gotten married and changed her name.

In the county with 25 high schools, the researcher received surveys from 58 teachers in 13 high schools. In the county with 17 high schools, the researcher received surveys from 52 teachers in 10 high schools. Based upon the State grading system, only three of the high schools were A-rated by the state. Six schools were rated as B schools, five were rated as C schools, seven were D-rated and one was rated as an F school. One school did not have a school grade because FCAT scores were available for only one academic year. Seven of the schools contained a student population of 50% or more that earned a proficient or higher score (3, 4 or 5) on the reading portion of the FCAT. Six

schools had 50% or more of their students enrolled in free and/or reduced lunch, and nine schools had minority student populations that were at or exceeded 50% of the total school population. Demographic information about the schools is provided in Table 2.

Table 2  
*Demographic Information for Schools of Teacher Participants*

	Number of Teacher Participants	School Grade Assigned by the State	% Students Scoring a 3, 4, or 5 on FCAT Reading	% Students on Free and/or Reduced Lunch	% Minority Students
School 1	1	P*	8	48	40
School 2	5	B	42	33	31
School 3	3	C	46	21	25
School 4	8	D	23	44	40
School 5	2	A	58	9	14
School 6	5	D	49	28	58
School 7	8	C	42	28	58
School 8	10	D	32	39	32
School 9	8	B	76	9	16
School 10	2	A	57	13	11
School 11	11	D	35	53	74
School 12	7	B	56	18	35
School 13	7	D	30	51	58
School 14	5	C	39	39	50
School 15	2	B	52	25	43
School 16	6	C	35	59	84
School 17	3	F	25	71	70
School 18	4	D	25	66	81
School 19	1	D	27	63	85
School 20	2	B	63	10	25
School 21	2	B	40	39	39
School 22	5	A	50	26	43
School 23	3	C	48	45	49

*Note:* This information is based upon information from the 2006-2007 school year. \*P means that the school grade is Pending.

Teacher data were also analyzed to see if there were any findings of interest between the teacher demographic variables and the teachers' self-efficacy scores. First, correlations were run between teacher years of teaching experience and efficacy and

teacher age and efficacy. The findings, shown in Table 3, indicate that the magnitude of these correlations was very small and none of the findings was statistically significant ( $p < .05$ ), but it is interesting to note that as teachers advanced in age and in years of experience, there was a negative correlation with teacher efficacy for student engagement. This was not the case with any of the other efficacy scales.

Table 3  
*Correlations between Teacher Efficacy Scores and Years Teaching and Age*

	Teacher Efficacy for Student Engagement	Teacher Efficacy for Classroom Management	Teacher Efficacy for Instructional Strategies	Total Efficacy
Years Teaching	-.02	.12	.16	.10
Age	-.06	.07	.10	.04

*Note:* Years Teaching: number of years of teaching experience.

Age: 1 = 21-25; 2=26-30; 3=31-35; 4=36-40; 5=41-45; 6=46-50; 7=51 and older.

With regard to the remaining teacher variables, it seemed more practical to investigate these more closely by examining mean efficacy rather than attempting to correlate ordinal and dichotomous variables with continuous variables. An analysis of mean efficacy scores by teacher degree achieved indicated that teachers with master's degrees and bachelor's degrees differed little in their self-perception of efficacy for classroom management and instructional strategies, but that teachers with master's degrees felt less efficacious with regard to student engagement. There was also a larger standard deviation among these scores for master's degree teachers. Overall, however, teachers with bachelor's degrees felt more efficacy than those with master's degrees. Little can be inferred about teachers with doctorates as the sample only included one doctorate. Mean scores by degree achieved are found in Table 4.

Table 4  
*Efficacy Score Means by Degree Achieved*

	Mean	SD
Teacher Efficacy for Student Engagement		
Bachelor's degree	52.73	8.31
Master's degree	50.31	10.79
Doctorate degree	55.00	-
Teacher Efficacy for Classroom Management		
Bachelor's degree	59.82	8.00
Master's degree	59.92	7.31
Doctorate degree	53.00	-
Teacher Efficacy for Instructional Strategies		
Bachelor's degree	57.86	7.33
Master's degree	57.67	8.60
Doctorate degree	61.00	-
Total Teacher Efficacy		
Bachelor's degree	170.41	20.80
Master's degree	167.89	23.22
Doctorate degree	169.00	-

*Note:* Bachelor's  $n = 73$ ; Master's  $n = 36$ ; Doctorate  $n = 1$ .

On average, teachers who were certified by the National Board for Professional Teaching Standards in addition to carrying a state certification felt more efficacious than teachers with only state certification. It is important to note, however, that there were only nine Nationally Board Certified teachers in the sample. See Table 5 for the comparison of efficacy scores.

Table 5  
*Efficacy Score Means by Type of Certification*

	Mean	SD
Teacher Efficacy for Student Engagement		
State Certification	51.76	9.30
Plus National Board Certification	54.11	8.01
Teacher Efficacy for Classroom Management		
State Certification	59.58	7.72
Plus National Board Certification	62.11	8.04
Teacher Efficacy for Instructional Strategies		
State Certification	57.63	7.54
Plus National Board Certification	60.00	9.67
Total Teacher Efficacy		
State Certification	168.98	21.30
Plus National Board Certification	176.22	23.41

*Note:* State Certification  $n = 101$ ; Plus National Board Certification  $n = 9$ .

Female teachers in the sample appear to feel more efficacious than male teachers in all four categories of efficacy, and the standard deviation around their scores is smaller. Again, however, it is necessary to remember the disparity in the sample with 95 female teachers and only 15 male teachers. Table 6 displays the mean efficacy scores by gender.

Table 6  
*Efficacy Score Means by Gender*

	Mean	SD
Teacher Efficacy for Student Engagement		
Male	51.27	11.02
Female	52.06	8.93
Teacher Efficacy for Classroom Management		
Male	58.07	8.86
Female	60.06	7.56
Teacher Efficacy for Instructional Strategies		
Male	54.60	9.52
Female	58.34	7.31
Total Teacher Efficacy		
Male	163.93	27.07
Female	170.46	20.46

*Note:* Male  $n = 15$ ; Female  $n = 95$ .



The sample of teachers was predominantly White (85%), so it is difficult to make any inferences about race and efficacy because so few non-White teachers responded to the surveys. Efficacy mean scores are displayed by teacher ethnicity in Table 7.

Table 7  
*Efficacy Score Means by Ethnicity*

	Mean	SD
Teacher Efficacy for Student Engagement		
Black	51.63	8.28
Hispanic	54.17	13.45
White	51.70	9.07
Other	58.50	6.36
Teacher Efficacy for Classroom Management		
Black	61.88	6.03
Hispanic	58.50	9.89
White	59.76	7.83
Other	57.00	4.24
Teacher Efficacy for Instructional Strategies		
Black	55.13	7.83
Hispanic	57.83	9.53
White	57.98	7.68
Other	61.50	3.54
Total Teacher Efficacy		
Black	168.63	20.94
Hispanic	170.50	31.63
White	169.44	21.19
Other	177.00	14.14

*Note:* Black  $n = 8$ ; Hispanic  $n = 6$ ; White  $n = 94$ ; Other  $n = 2$ .

*Students.* All information for students in this study was provided to the researcher by each county's Office of Accountability and Research without any identifying features such as student names or student identification numbers. Students were selected for inclusion in the study when one of the teachers responding to the surveys identified the course section identification number of one ninth or tenth grade English class that he or she taught. The teacher was asked to contemplate the character of this one identified class

while completing the survey. Data for students in those section identification number classes were obtained from the Offices of Accountability and Research and matched to each participating teacher by the course section identification number. Original data provided by the county included 2,129 students. The ethnic make-up of the original sample was 57 Asians, 488 Blacks, 441 Hispanics, 1133 Whites, and 10 Others (multi-racial or Eastern Indian). Because the population of Asian and multi-racial or Eastern Indian students was so small, these students were eliminated from the sample, leaving a sample size of 2,062. FCAT reading developmental scale scores can only fall in the range of 86 to 3008 points, and one score fell outside of this range. It was considered to be a data entry error and eliminated from the study, leaving a total student sample size of 2,061.

On the Teacher Personal Information Questionnaire (Appendix D) distributed to teachers, question 9 stated:

Efficacy is believed to be context specific, so as you respond to items on the Teachers' Sense of Efficacy Scale, keep one class of ninth or tenth graders whom you currently teach English I or II in mind. Circle the period and class title of that particular class in the list above [where teachers outlined their daily teaching schedule] and describe the character of the selected English I or II class.

Fifty-three of the 115 teachers responded to this question, and this qualitative information provided some insight into the types of students and classes included in the sample. In a review of responses, comments seemed to fall into six categories. Some responses gave purely demographic information, some explained the varying level of abilities present in the class, others highlighted the motivation or lack thereof in the class and a few hinted at

positive classroom transitions. There was also a set of miscellaneous comments that defied classification. The question was phrased in such a manner that teachers could interpret it in numerous ways. This was deliberate on the part of the researcher because the question was an attempt to ascertain what it was about this particular class that stood out in the responding teacher's mind.

Many teachers gave purely demographic information. This is a reasonable response, given the fact that all of the previous questions on the Teacher Personal Information Questionnaire had asked for demographic information about the teachers themselves such as years teaching experience, highest degree achieved, age, ethnicity and teaching schedule. These are the responses of teachers who gave mainly demographic information about their classes:

- In this class I have a great blend with my 34 students in cultural ethnicity, aptitude, and attitude. 23 males and 11 females in the class. (English I)
- Good general level class: 17 boys and 6 girls. (English I)
- My class is typically made up of white magnet students. I have several more girls than boys. (English I Honors)
- 76% minorities and first time honors students. AP course are targeted for my students. No "extreme" behavior problems. Two-thirds are magnet students. (English II Honors)
- Mixed gender, race and magnet/ non-magnet. (English II)
- Varied ability levels, ethnic make-up, and motivation. (English I Honors)
- 10 males, 5 females. 3 Hispanic, 12 black. 3 magnets, 12 traditional. Mostly low socioeconomic. Three ESE. (English I)
- Traditional students, majority African-American; even distribution of males and females. (English I)
- Diverse – 1/3 African- American, 1/3 Hispanic, 1/3 White. (English I Honors)
- It is a traditional class of 25 students. 10 students are in credit recovery, so they are repeating the tenth grade. It is at the end of the day, so the students are easily distracted from learning. (English II)
- This class is comprised primarily of students reading below grade level on the FCAT. (English II)
- This is a tenth grade regular class. The student is generally level 2 and 3 on FCAT reading. A good class. (English II)

Other teachers seemed to focus on the varying abilities of the students in their classes:

- This is an English II regular class with 35 students among whom 2 are seniors. Very different levels, some students are somewhat disruptive. 2 students are ESE (low level). (English II)
- Overall the class runs very smoothly. This class has a great diversity of students and participation is good as well. The biggest challenge is trying to come up with lesson plans that are adequate for all levels of students. (English II)
- Varying levels of readers which makes for group work and creative projects. Generally respectful students and, while not exactly brilliant, have the capacity and desire to succeed. (English I)
- I have 19 students with varying abilities - mostly level 1 and 2 readers but a few 3 and 4 readers. They have little behavior problems. (English I)
- My second period class has large range of abilities and performance. Seventeen students scored a 2 on the ninth grade reading FCAT. I even had two who scored a one. Students are often too sociable – taking time to settle even with bell work. (English II Honors)
- The character of this honors class is mixed, some belong and some don't (academically worthy). Most are needy and willing to learn. (English I Honors)
- The class is made up of a mix of students who are from various socioeconomic groups. A handful of students are not engaged and interested on a daily basis and require special attention. The majority of the group, however, is motivated and classroom management is not a problem. (English II)
- Most of the students are doing very well, but the others belong in regular English. (English I Honors)
- The English I class has many hard workers. Most of the students are dedicated and serious. However, there are a few that do not seem to be concerned with their academic performance. (English I Honors)

The majority of teachers (53%) that responded to this question, however, concentrated on their students' motivation or lack thereof to participate in class by completing assignments and staying on task. Thirteen teachers expressed frustration with their students in this regard:

- Low motivation, but less behavior problems than reading class. (English II)
- Unruly brats. (English I Honors)

- 1/3 attentive, 1/3 moderately attentive, 1/3 disinterested. Numerous unexcused absences - rarely make up work even with opportunities provided. Some strides made with students who performed poorly last semester, but only a few. Homework is done on an inconsistent basis. Students have difficulty coming to class prepared – pen, paper, test. Little parental feedback. Good kids at heart, but extremely behind in basic skills -- heading paper with full names, etc. (English I)
- Somewhat lazy – unmotivated. (English I)
- Large “regular” English class with one ESOL student and six ESE students, one of whom is ESOL. I have a few students who are consistently disruptive, but I have four or five who will (try to) sleep whenever we read – they participate when we are doing other activities (e.g. grammar, vocabulary, etc.). The reading level of this class is about half FCAT 1 and half FCAT 2. (English II)
- The question is not clear – character as I interpret it – My class is regular students with average to below average abilities and drive. (English I)
- This class is made up primarily of minority students. There are approximately 21 student in the class. The class is generally quiet and often has to be prodded to speak and participate. (English I Honors)
- English 2 regular, period 4: mostly boys and 75% are failing if not more. Most of them choose not to do assignments and, therefore, try to distract others. High absence rate, too. (English II)
- Of the 25 student I have, about 7 really struggle in reading comprehension. They’ve learned that I don’t tolerate disrespect or talking, so they’re very good. They beg every day not to do work – but we do. Because English is a second language to some of them, we have to work hard on reading comprehension and writing. (English I)
- My second period English II class is a fuse class with 11 ESE students. This is a very low level class with below basic skill level. They are very rambunctious and have no desire to succeed in this class. They are rarely prepared and it is even rarer for them to turn in homework. (English II)
- Period 5 is at a challenging group, but it is also a group with much talent and potential. The students are generally able to do grade level course work, but are often not inclined to complete class assignments, do homework, etc. The kids are usually friendly and talkative, sometimes a little too much so. (English I)
- Mostly remedial class: four natural freshmen, eleven repeating students who should be in grades 10-12. All natural freshmen are ESE students. Students are not particularly motivated by or interested in school, but they enjoy group work. We often modify assignments so they don’t require a homework component. Since it is a small class, we know the students very well and gentle teasing works wonders! The class is less structured in terms of assigned seats, raising hands, etc. The class is ethnically diverse. (English I)
- Diverse, impatient, easily thrown off task, disruptive. (English II)

Fifteen teachers, however, praised their students' motivation and work ethic:

- English I class is well behaved. They very rarely have discipline issues. We have many children that have fun with the literature as well as work hard. (English I)
- Very upbeat, all magnet kids, mostly females, mostly white students. (English I Honors)
- The class overall has about 70% of students who are eager to know what is expected of them and strive to meet those expectation. (English II)
- Very good students, not totally against being in class and learning. They are able to stay with me during a class discussion. They answer questions, they volunteer to read, and they interact with each other. Not all of them are on “grade level,” but they help each other out! (English I)
- These are, for the most part, diligent workers who are, again, for the most part, highly motivated learners. My students are respectful and have a good work ethic. (English I Honors)
- Very well-behaved but sluggish / lazy first thing in the morning! Most students have not passed the FCAT and are level 1 readers. They seem to value education; I consider them respectful and caring overall. (English II)
- Rather immature, but motivated. The students need help with fluency and comprehension. (English I Honors)
- Animated! Eager! Smart! (English II Honors)
- They are a fantastic group of students with vitality and energy. Mostly complete work on time and are quick to answer questions and get involved. (English I Honors)
- This class is a heterogeneous class with levels ranging from honors level students to low functioning students. The students in this class are mostly hardworking, quiet, and respectful. Most show a desire to learn, although there are a handful that have very poor attendance. They work hard when they are present, but they rarely are present! (English I)
- Large group of student that require additional encouragement to begin work at the bell. Once started, the class produces the most insightful, interesting discussions and usually submits assignments over and above minimum requirements. (English I)
- A large hyper class. Overall bright, the most true honor class I have taught in years, but a handful who don't belong. Chatty and witty. (English I Honors)
- My third period is comprised of 21 high achieving and motivated students. They are generally eager to learn and enjoy writing and literature. (English II Honors)
- These students are fun to teach! They are, for the most part, very enthusiastic and willing to work on all the material we are covering. They ask great questions, and many of them respond with insightful answers. (English I Honors).
- This class is a mix of students, most of whom should be in an Honors class, with a few others who should be in regular class. They are usually willing to put in the effort required for the level of work expected of the class. (English I Honors)

Some teachers' answers defied categorization:

- Young enough to care what I think but old enough to have some very poor work habits. (English II)
- They are lazy, jovial and typical teenage students. I usually have to spend time throughout the class period teaching my students manners and respect. (English I)

And a few teachers' responses expressed hope that students' attitudes were changing for the better throughout the course of the year:

- English I is a world literature course contains 18 students ages 14-16. The ability/skill level is extremely low. Most students entered my class in January claiming to have never read an entire book. At the beginning of the semester, six of these students arrived as hold backs from middle school. Many entered with very negative attitudes toward school, especially English. Fourteen speak Spanish as their first language. Of the remaining four, two are African-American and two are Caucasian. Over the past two months, the class climate has changed and become much more positive. (English I)
- I have 26 students who range from level 1-4, most being on the lower end of that scale. They are very unmotivated to read and several fail the course. I have seen truly amazing results with an outside reading unit I just completed, however. The books captured students, and the use of literature circles prompted intelligent discussion. This is the one activity all year that I got students to buy into. They all purchased their own books and most have continued to read books by the same author. (English II)

The fact that many teachers provided purely demographic data was not surprising since the request to "describe the character of the class" followed numerous questions about teacher demographic information. What is surprising, however, is the variety in the responses. The researcher had assumed that most teachers would choose to concentrate on their best class, the brightest or the most well-behaved, because the class' scores would probably reflect most positively upon the teacher. From the responses, though, it is obvious that not all teachers reported their perceived self-efficacy with their best class in mind. Descriptions ranged from "unruly brats" to "Animated! Eager! Smart!"

### *3.3.2 Data Collection*

Teacher data were collected from teachers at the school sites that agreed to participate in the study. Student achievement and demographic data regarding students' ethnicity, socioeconomic status, grade and test scores were obtained from each county's Office of Accountability and Research.

The researcher gathered the information from the teacher surveys with the assistance of the Language Arts/English supervisors for each county and the language arts department chairs. At the invitation of the Language Arts/English supervisor for each county, the researcher attended one of the monthly meetings of language arts department chairs. At these meetings, the researcher explained the research (script can be found in Appendix A) and provided each department chair with a packet containing directions for administering the survey (Appendix B) to the English teachers in their department as well as letters to each teacher about the research (Appendix C), an adequate supply of Teacher Personal Information Questionnaires (Appendix D) and OSTES surveys (Appendix E) as well as Informed Consent Forms (Appendix F). Blank envelopes were also provided, so teacher surveys would be kept confidential. Department chairs were asked to administer the surveys at their next departmental meeting. Specifically, department chairs were asked to explain the research, request that their teachers fill out the surveys and return them to the department chair sealed in the envelopes provided. Each department chair who participated returned the surveys to the researcher in a postage paid envelope that was also provided.



It is important to note that teachers filling out the survey were asked to identify one specific class of ninth or tenth graders that they were teaching at the time of the survey, by course section identification number in one county and by class period in the other county, and to think about that one specific class while scoring themselves on the OSTES. The rationale for making this request is that teacher efficacy is both context and subject matter specific (Pintrich & Schunk, 1996; Bandura, 1997). Therefore, by asking teachers to keep one particular class in mind, the researcher could utilize the course section identification number or class period to match each teacher with the particular class of students that he or she had in mind while responding to the survey.

After the surveys were returned to the researcher, the researcher worked with each county's Office of Accountability and Research to obtain demographic and test score information for students. Because the responding teachers had indicated a specific class on their surveys, the researcher was able to provide the course section identification number or the teacher name and period, depending on the county's record keeping system, to each Office of Accountability and Research to obtain student information while still keeping student names and identification numbers confidential.

### *3.3.3 Data Management.*

For security purposes, all electronic files associated with and generated from the data (e.g., SAS programs and output) were encrypted and stored on a password protected external hard drive that was kept in a locked file cabinet when not in use. The researcher was the only person who knew the password to access the encrypted files. Similarly, the researcher's laptop, which was used to conduct the data analysis, was password protected

and programmed to lock after 10 minutes of inactivity. Only the researcher knew the password to unlock the computer. Raw data in the form of Teacher Personal Information Questionnaires and OSTES surveys were kept in a locked file cabinet and will be shredded at the conclusion of the investigation.

Prior to conducting any analysis, several data management tasks were completed. First, each teacher was assigned a participant number to protect his or her privacy. Teachers' schools were also assigned a code for anonymity purposes. Second, data were checked to ensure that the researcher had the proper course section identification numbers for each teacher. Data were also checked to make certain that each teacher was indeed a teacher of grades 9 or 10 English and that FCAT scores fell within the acceptable range of 86 to 3008. Finally, to improve data analysis processing time, a smaller data set that contained only methodological variables (e.g., demographic data) and substantive variables of interest (e.g., dependent and independent variables) was created from the data files provided by each district.

#### *3.3.4 Dependent Variable*

*Student Reading Achievement Gain.* The created variable of student reading achievement was used as the dependent variable in the study. The variable was created by subtracting the 2006 FCAT developmental scale scores (DSS) from the 2007 FCAT DSSs. Creating this variable allowed the researcher to control for student prior achievement. It is important to note that developmental scale scores mean different things at different grade levels. In FCAT reading, scale scores, ranging from 100 to 500 for each grade level, are converted to developmental scale scores (DSS or vertical scale scores),

which place the scores of students on a scale ranging from 86 to 3008. This continuous scale allows student progress to be tracked from year to year. In order to create developmental scale scores, anchor items are placed on tests given in different years to relate the scores of the current year to the scores of the first year of operational testing. Hence, the developmental scale score is based on linking items, items that appear identically on the tests of the adjacent grade levels, to relate the scores from one grade to those in the grade above and the grade below. Appendix J provides a detailed explanation of how DSS are created. The explanation is taken directly from a publication entitled *Reading and Mathematics: Technical Report for 2005 FCAT Test Administration* (Human Resources Research Organization, 2006).

With the scale score from each grade level linked to those above and below it, a single scale is created. The intended use of the DSS is to monitor the progress of individual students over time. By comparing a student's scores in the same FCAT subject for two or more years, it is possible to determine if a student's performance improved, declined or remained constant. Developmental scale scores were used in this study because the scores and the performance level cut points are constant from year to year; those used for Sunshine State Standard scale scores change from year to year. Developmental scale scores cannot be used to compare across grades, however. For instance, a ninth grader's DSS cannot be compared to a tenth grader's because the two are operating on different scales. Because of this, the student variable of grade level was added as a covariate in the research design. Performance level cut points on developmental scale scores for ninth and tenth grade students are shown in Table 8.

Table 8

*Reading Developmental Scale Scores (86 to 3008) for each Achievement Level on the Florida Comprehensive Assessment Test for Ninth and Tenth Grade*

	Level 1	Level 2	Level 3	Level 4	Level 5
Grade 9	772-1771	1772-1971	1972-2145	2146-2297	2298-2943
Grade 10	844-1851	1852-2067	2068-2218	2219-2310	2311-3008

*Measurement Instrument for the Dependent Variable.* The FCAT is a statewide assessment that consists of various tests that measure achievement in reading, writing, mathematics and science. The test includes a criterion-reference test based upon the Florida Sunshine State Standards (SSS) and a norm-referenced test (Florida Department of Education, *Understanding FCAT Reports*, 2006).

The FCAT Reading SSS test includes a wide variety of written material to assess students' reading comprehension. It includes both informational and literary reading passages. Informational passages consist of magazine and newspaper articles, editorials and biographies. Literary passages include short stories, poems, folk tales and selections from novels. The reading portion of the test is comprised of traditional multiple-choice items as well as items requiring students to compose short and extended responses.

Technical information about the FCAT was provided by the most current report available, the *Technical report: For operational test administrators of the 2000 Florida Comprehensive Assessment Test* (Florida Department of Education [FLDOE], 2001). The report describes the purpose of FCAT administration and provides its psychometric properties. Traditional Cronbach's alpha reliability estimates are given for both the Grade

8 reading (.82 for the Informational Text and .79 for the Literary Text) as well as the Grade 10 reading (.80 for the Informational Text and .68 for the Literary Text). Marginal reliability scores, scores that indicate whether or not the FCAT has reliabilities similar to those of other standardized and statewide tests, are included as well. The marginal reliability for the grade 8 reading test is .91, and the marginal reliability for grade 10 is .88. The lower reliabilities of the Cronbach's alpha compared to the marginal reliabilities reflect the fewer number of test items on the subtests of the FCAT.

### *3.3.5 Independent Variables*

*Teacher Efficacy.* The independent variables for this study included a total score for efficacy as well as three subscale scores for efficacy in the following areas: efficacy for instructional strategies, classroom management and student engagement.

*Measurement Instrument for the Independent Variables.* The Ohio State Teacher Efficacy Scale (OSTES) was used to obtain measures of teacher efficacy. The survey contains 24 items in a nine-point Likert format anchored from a score of 1 indicating "nothing" to a score of 9 indicating "a great deal." In addition to a total score for efficacy, the measure can be further broken down into three subscales: efficacy for instructional strategies, classroom management and student engagement. Reliability for the teacher efficacy subscales are 0.91 for instruction, 0.90 for management and 0.87 for student engagement. Construct validity for the survey was determined by assessing the correlation of the OSTES and other existing measures of teacher efficacy. Positive correlations (with the two-question RAND scale:  $r = 0.18$  and  $0.53$ ,  $p < 0.01$ ; with Gibson and Dembo's Teacher Efficacy Scale:  $r = 0.64$ ,  $p < 0.01$ ) provide evidence for

construct validity. Chapter 2 includes a detailed explanation of the development of the OSTES.

Because the OSTES is a relatively new measure of teacher efficacy, the researcher conducted a confirmatory factor analysis (CFA) to evaluate construct validity using Mplus version 3.0 (Muthen & Muthen, 1998-2004). Before conducting the CFA, however, the response distributions for the 24 OSTES items were examined for univariate normality using SAS v9.1.3. To date, only one CFA has been undertaken by Roberts and Henson (2001), and this CFA was done while the instrument was still under development. As mentioned in Chapter 2, Roberts and Henson (2001) found support for the factors of efficacy for student engagement and instructional strategies, but they found the classroom management factor to be weak and recommended eliminating it from the instrument. Because of this, Tschannen-Moran and Woolfolk Hoy (2001) modified the instrument.

Thus far, no outside factor analysis has been conducted on the modified scale. Confirmatory factor analysis is appropriate in this case because exploratory factor analysis is considered to be a theory-generating procedure while CFA is a theory-testing procedure that is fitting when there is a strong theoretical and/or empirical base (Stevens, 2002). Gorsuch (1983) stated that exploratory factor analysis should be “reserved for those areas that are truly exploratory, that is, areas where no prior analysis have been conducted” (p. 134). The work done by Tschannen-Moran and Woolfolk Hoy (2001) to validate the OSTES provides a strong empirical base. In a CFA, the researcher specifies the structure of the factor models a priori. According to the analysis conducted by

Tschannen-Moran and Woolfolk Hoy (2001), there are three factors that emerge from the data: teacher efficacy for student engagement, instructional strategies and classroom management. In a CFA, the researcher proposes competing models based upon the data available by “fixing” or “freeing” specific parameters such as the factor correlation coefficients, the factor coefficients or the variance/covariance of the error measurement.

According to Gillaspay (1996):

Fixing a parameter refers to setting the parameter at a specific value based on one’s expectations. Thus, in fixing a parameter the researcher does not allow that parameter to be estimated in the analysis. Freeing a parameter refers to allowing the parameter to be estimated during the analysis by fitting the model to the data according to some theory about the data. The competing models or hypothesis about the structure of the data are then tested against one another. (p.7)

Once the analysis is completed, several different statistics are produced by the computer software for determining the fit of the competing models. The acceptability of the fitted CFA solution in this study was evaluated on the basis of three aspects: (1) overall goodness-of-fit; (2) the presence or absence of localized areas of strain in the solution, determined by assessing the standardized residuals and the modification indices; and (3) the interpretability, size, and statistical significance of the model’s parameter estimates. Campbell, Gillaspay and Thompson (1996) recommend using several indexes of fit to determine relative fit.

### *3.3.6 Contextual Variables Taken Into Consideration*

Throughout the United States, there is a call to raise student academic achievement. Simultaneously, there is a growing movement to hold school administrators

and teachers accountable for the quality of education they deliver. One message is loud and clear: variables among student populations, such as SES, ethnicity and gender, can no longer be used as an excuse. While these variables should not be used as an excuse, Ballou, Sanders, and Wright (2003) exhort that neither can they be ignored:

Although there is evidence that virtually all students, even those from disadvantaged backgrounds, can succeed in the right educational environment, it is much more difficult to raise the achievement of disadvantaged children to the new standards. Holding teachers and administrators accountable for student outcomes without regard for the differences in students' backgrounds is manifestly unfair and, in the long run, counter-productive. Such policies will alienate educators, making it more difficult to staff schools serving the neediest population. The perception that educators are being held accountable for student achievement without due regard for factors beyond their control may ultimately discredit the standards movement itself. (p. 37)

One solution to this dilemma has been to take an alternative approach and measure for student progress rather than overall achievement. This is generally done by controlling for students' initial levels of achievement either through utilizing pre- and post-test scores or by introducing a prior score on the same test as a regressor to control for prior achievement. For instance, in this study, the dependent variable was the students' reading achievement gain, a variable created by subtracting the 2006 FCAT DSS from the 2007 FCAT DSS. Creation of the reading achievement variable enabled the researcher to take prior achievement into consideration. Additionally, because this method measures gains from the student's own starting point, Ballou, Sanders, and Wright (2003) note that many contend that it also "implicitly controls for socio-economic status and other background factors" (38).



Such methodology comes under criticism, however, by those who believe that factors such as SES and other demographic factors affect not only a student's starting point but also the rate at which they learn (Linn, 2001; Kupermintz, 2002). Studies by the Value-Added Research Consortium at the University of Florida College of Medicine estimated numerous value-added models with and without contextual variables such as socio-economic status and ethnicity. They found that these variables were almost always statistically significant and that teacher and school effects were sensitive to the presence of these variables as controls (University of Florida, 2000a, 2000b). Members of the research community are skeptical that the contribution of school and teacher level factors to student achievement can accurately be measured without controlling for such contextual variables (Darling-Hammond, 1997). For this reason, the contextual variables of SES, grade and ethnicity were controlled for in this study.

Even if researchers do control for student contextual variables such as socio-economic status and ethnicity, however, how is one to know that there are not other factors influencing student achievement? Inclusion of student-level variables does not account for claims that disadvantaged students are frequently and systematically assigned to less effective schools that employ less effective teachers. For this reason, Aitkin and Longford (1986) state that “[e]very effort needs to be made in school comparisons to avoid model mis-specification by the inclusion of all relevant variables in the initial model” (p.22). With this in mind, the following contextual variables at the student level were taken into consideration:

*Ethnicity.* The ethnic affiliation of a person as reported by the school district to the testing agency: African-American, White, Hispanic or Other. Originally, all four categories were controlled for, but it was determined during the analysis that models would be more parsimonious if the variable of ethnicity was reduced to White and non-White. Furthermore, in original analysis of the data, Black and Hispanic students performed similarly on the FCAT, so it was determined that the variables could be combined without losing the unique contributions of each. The original data provided by the county included 2,129 students. The ethnic make-up of the original sample was 57 Asians, 488 Blacks, 441 Hispanics, 1133 Whites, and 10 Others (multi-racial or Eastern Indian). Because the population of Asian and multi-racial or Eastern Indian students was so small, these students were eliminated from the sample. One of the Hispanic student's test scores fell outside of the acceptable range. This was deemed to be a data entry error, and the student was eliminated from the sample. This left 488 Blacks, 440 Hispanics and 1133 Whites for a total sample of 2,061 with 928 non-White (45.03%) and 1133 White (54.97%).

*Gender.* The sex of a person as reported by the school district to the testing agency: male or female. Initially, gender was hypothesized to be an important variable because research indicates that females outscore males in tests of verbal ability. In the analysis, however, gender was determined to hold no statistical significance, so it was not factored in as a control variable. There were, however, 1020 females (49.49%) and 1041 males (50.51%) in the sample.

*Grade.* Because developmental scale scores mean different things at different grade levels, they cannot be used to compare across grades. For instance, a ninth grader's DSS cannot be compared to a tenth grader's because the two are operating on different scales. Because of this, the student variable of grade level was added as a covariate in the research design. There were 1123 ninth graders (54.49%) and 938 tenth graders (45.51%) in the sample.

*Prior reading achievement.* Prior reading achievement was taken into consideration by creating the dependent variable of reading achievement gain by subtracting the 2006 FCAT DSSs from the 2007 FCAT DSSs for reading. Doing so allowed the researcher to focus on reading gains rather than final test scores. The mean reading achievement gain for ninth graders was 52.30 (SD=166.70) and 32.59 (SD=187.98) for tenth graders.

*Socio-economic status.* This is an economic designation assigned to students determined by a student's participation in the free and reduced lunch program. For purposes of this study, SES is divided into two categories: those students who are eligible for the free or reduced lunch program and those students who are not eligible for free or reduced lunch. According to the website (School District of Hillsborough County, 2006) that explains eligibility guidelines, the National School Lunch and Breakfast program uses United States Department of Agriculture Income Eligibility Guidelines which are adjusted annually for inflation and are used when approving applications for children who apply to receive a free or reduced meal. The free lunch guidelines are 130% of Federal poverty guidelines, and the reduced lunch guidelines are 185% of poverty

guidelines. In the sample for this study, 1211 students (58.76%) of the students did not qualify or apply for free and/or reduced lunch while 850 students (41.24%) did qualify for free and/or reduced lunch.

### 3.4 Data analysis

This research investigated the degree to which teacher efficacy (a class-level factor) overall and in three subscale areas, instructional strategies, classroom management and student engagement, influenced student reading achievement gains (an individual-level outcome) after taking certain contextual variables with regard to students' SES, grade and ethnicity into consideration. Because student achievement occurs at the individual level and teacher efficacy influences students at the classroom level, the unit of analysis becomes problematic. Conventional methods require single-level analysis which leads to the conceptual and empirical problems associated with examining student-level and teacher-level variables. Multilevel modeling, on the other hand, avoids misestimation standard errors, aggregation bias and heterogeneity of regression problems that may affect the results of ordinary least squares regression analysis of data in which one or more individual-level characteristics are aggregated to the group level.

#### *3.4.1 Advantages of Hierarchical Linear Modeling (HLM)*

HLM is a multilevel modeling technique intended for nested data (Bryk & Raudenbush, 2002) because it models the data structure that is created in education settings with students nested within classrooms, nested within schools, nested within

counties. For the purposes of this study, however, the researcher focused on student-level variables and classroom-level variables.

HLM also takes into account the interdependence of individual measures collected within the same organizational unit, the classroom in this case. While classical statistical analyses are likely to be biased if the issue of independence is disregarded, HLM assesses student-level, classroom-level and the between-level effects simultaneously (Marsh, Kong & Hau, 2000). This is an important feature because it takes into account that students are not randomly placed into classes. Often, levels of ability affect course scheduling. For instance, certain English classes may be considered 'honors' level. Or, as is often the case, because advanced mathematics classes may only be offered a few periods a day, those students will all end up in the same English class simply because of scheduling issues.

Other problems with classical statistical analysis are related to the levels of analysis and aggregation bias. Often, data can only be collected at one particular level, thereby limiting the inferences made to that particular level of analysis; however, bias is often created by making inappropriate inferences to other levels of the data. Aggregation bias is one of the most habitually committed errors in classical statistics (Hanushek, Rivkin & Taylor, 1997). HLM partitions the variance in a dependent variable into its within- and between-classroom components. Therefore, it is possible to model the effect of teacher efficacy on only the portion of the variance in student achievement occurring between classrooms while modeling the influence of students' SES, grade and ethnicity without aggregating these variables at the classroom level. In other words, it is possible to

predict values of the dependent variable based on a function of predictor variables at more than one level (Luke, 2004).

A final advantage of using HLM is related to sample size. In classical statistical analysis, sample number must meet specific criteria. In the real world classroom, however, some teachers may teach a class consisting of 14 students, while another teacher teaches a class comprising 28 students. HLM allows the researcher to include data from all classes regardless of class size (Bryk & Raudenbush, 2002).

Taking all of the factors into consideration, there is ample support for the use of HLM as the statistical analysis for this study. The first consideration is the nested structure of the data; students are grouped into classes. In addition, students are not randomly assigned to classes, and issues of interdependence must be acknowledged. The research questions also address questions that can only be answered by including two levels of analysis because of the inclusion of variables at both the student level and the class level. Furthermore, classes typically vary in size, and HLM can accommodate such data.

#### *3.4.2 Characteristics of HLM*

Hierarchical Linear Modeling (HLM) was designed for data consisting of nested units of analysis. In the present study, level-1 is the smallest unit of analysis, the students. In this first level, the data includes the contextual student variables of SES, ethnicity and grade as well as prior student achievement through the creation of the dependent variable of reading achievement gain. Level-2 represents the grouping variable; in this study, it is the classroom led by the teachers with varying levels of efficacy. A large amount of

variation within and between these levels can be present, and Bryk and Raudenbush (1992) posit that 80-90% of the within-group variation is lost when aggregating student characteristics within classes. HLM, on the other hand, allows for the variance to be included.

The basic principle of HLM is comparable to that of Ordinary Least Squares (OLS) regression. At the student-level (level-1), an outcome variable is predicted as a function of a linear combination of one or more level-1 variables:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_1 + \dots + r_{ij}$$

where:  $\beta_{0j}$  represents the intercept of the group,  
 $\beta_{1j}$  represents the slope of variable  $X_1$  of group  $j$ , and  
 $r_{ij}$  represents the residual for individual  $i$  within group  $j$ .

On subsequent levels, the level-1 slope(s) and intercept become dependent variables being predicted from level-2 variables, as shown in the following models:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_1 + \dots + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_1 + \dots + u_{1j}$$

where  $\gamma_{00}$  and  $\gamma_{10}$  are the intercepts and  $\gamma_{01}$  and  $\gamma_{11}$  are the slopes predicting  $\beta_{0j}$  and  $\beta_{1j}$ , respectively from variable  $w_1$ . Through this process, the effects of level-1 and level-2 variables can be modeled to reflect the organizational structure of the data. By predicting the slopes as well as the intercepts (means), an attempt is made to explain the differences in the relationship between level-1 and level-2 (Luke, 2004).

### 3.4.3 Methodological Points Related to the Use of HLM

There are some issues with HLM that were considered prior to building models. The researcher considered sample size and effect size in order to determine power as well as determine what method of centering would best suit the data. In addition, the researcher decided whether or not slopes and intercepts would be allowed to vary. These issues will now be discussed.

*Power, effect size and sample size.* Power is the probability of rejecting the null hypothesis when a specific alternative hypothesis is true. In other words, it is the chance of making a correct decision. Power, however, depends on many factors: within-cluster sample sizes (or  $n$ , the number of students in each class), the number of clusters (or  $N$ , the number of classes), the inter-class correlation (or the ICC, the variability between clusters) and the desired effect size (or the population means difference of the two groups divided by the standardized error of the outcome), while holding the significance level constant (Spybrook, Raudenbush, Liu & Congdon, 2006). Power is also affected by the number of contextual variables being taken into consideration. Literature on sample size in HLM (Snijders & Bosker, 1992; Kreft, 1992; Raudenbush & Liu, 2000; Bryk & Raudenbush, 1992) is inconclusive. In order to achieve a desired level of power at a hypothesized effect size and a chosen level of significance, researchers describe various methods of selecting their sample sizes.

Snijders and Bosker (1992) created the formula  $N(n)$  where  $N$  = the number of classes and  $n$  = the number of students in the class to determine the sample size. They recommend that the number of classes ( $N$ ) be greater than or equal to 30 and that the



number of students per class ( $n$ ) be less than 28 to keep the standard error below the desired level. The standard error for cross-level interaction can be kept low if the number of classes ( $N$ ) stays between 30 and 62 and the number of students per class ( $n$ ) ranges from 11 to 28. According to Snijders and Bosker (1992), maximum power for the test of cross-level interaction can be obtained by choosing 42 classes each with 19 students per class, but, at the second-level, maximum power can be achieved by choosing the largest number of classes ( $N$ ) possible. They deduce that researchers should try to sample as many classes ( $N$ ) as possible and that the number of students per class should be at least greater than 10 although greater than 25 is most desirable. Kreft (1992), however, disagrees with the need for so many students per class and posits that groups can consist of only one observation as long as the data includes at least two groups with multiple observations.

Utilizing Cohen's (1988) standardized effect size and adding a standardized measure of site-by-treatment variance as well as a standardized measure of site-level moderating effects, Raudenbush and Liu (2000) outline another method for determining power in HLM that calls for careful consideration of the number of individuals sampled at each site and the number of sites. While both of these factors are important in contributing to power, the number of groups is more important.

Power for this study was determined *a priori* utilizing Optimal Design software (Spybrook, Raudenbush, Liu & Congdon, 2006). In this software package, the researcher entered the significance level at which to test ( $\alpha = .05$ ), the estimated number of units per cluster, or students in each class, ( $n = 25$ ) and the effect size ( $\delta = .20$ ). The program

created a graph to determine the number of clusters (J) that the researcher had to have in order to achieve the desired power. The desired power in this study was .80, so a total of 71 clusters, or teachers, was needed to achieve the desired power.

*Intercepts- and Slopes-as-Outcomes.* The decision of whether to create models that allow intercepts and slopes to vary depends upon whether or not there is significant variation among classrooms. The intercept parameters represent the expected value of outcome measure when all of the predictor variables are zero; the slope parameter represents the expected increase in the outcome variable for every unit of change in the predictor variable (Hoffman, 1997). The researcher has four modeling options with regard to slopes. Slopes may be fixed, random, non-random varying or the researcher can opt to use the slopes-as-outcomes model. In a fixed model, the effect of the level-1 variable is held constant across all level-2 units:

$$B_{2j} = \gamma_{20}$$

In a model with random slopes, the slopes are allowed to vary as a function of the level-2 slope plus the random effect ( $u_{2j}$ ) associated with each level-2 unit:

$$B_{2j} = \gamma_{20} + u_{2j}$$

In a slopes-as-outcomes model, part of the variation of the slope is predicted by the level-2 variable ( $W_j$ ), and a random component ( $u_{2j}$ ) remains unexplained:

$$B_{2j} = \gamma_{20} + \gamma_{21}W_j + u_{2j}$$

A non-random varying slope is preferable in instances when, once the effect of the level-2 variable ( $W_j$ ) is taken into account, the effect of the residual variation ( $u_{2j}$ ) is

immaterial. In a non-randomly varying slope model, the random component ( $u_{2j}$ ) is removed, and the slope varies only as a function of the predictor variable ( $W_j$ ) at level-2:

$$B_{2j} = \gamma_{20} + \gamma_{21}W_j$$

According to Bryk and Raudenbush (1992), modeling random variation in a slope that does not vary among groups is statistically inefficient. The models created for the analysis of the research questions had fixed slopes because the researcher was interested in seeing the impact of the classroom teacher's efficacy on his or her students alone, not the impact of one teacher's efficacy across classrooms.

#### *3.4.4 The HLM Models*

The intent of this study was to investigate the effects of teacher efficacy for instructional strategies, classroom management and student engagement on ninth and tenth grade students' reading achievement. Teachers with high efficacy in these three areas were hypothesized to positively impact students' reading gains. Student contextual variables of prior student achievement, socioeconomic status (SES), ethnicity and grade were controlled for in the study. At this point, a reminder of the research questions may be necessary:

While taking student characteristics (prior student achievement, SES, gender and ethnicity) into consideration:

1. What is the influence of teacher efficacy regarding instructional strategies on student reading achievement gains?
2. What is the influence of teacher efficacy regarding classroom management on student reading achievement gains?

3. What is the influence of teacher efficacy regarding student motivation on student reading achievement gains?
4. What is the influence of total teacher efficacy (teacher efficacy regarding instructional strategies, classroom management and student motivation combined) on student reading achievement gains?

The research questions were examined using hierarchical linear models with individual students nested in teacher classrooms. All multivariate analyses were conducted using PROC MIXED in SAS v9.1.3 (SAS Institute Inc., 2003) However, before conducting any multivariate analyses, data were screened for violations of assumptions often associated with multilevel models (i.e., multicollinearity, normality, linearity and homogeneity of variance). Further, the data screening techniques described below are the same as those recommended by Hox (2002) and Raudenbush and Bryk (2002).

First, the data were examined for multicollinearity. In addition to the bivariate examination of independent variables via zero-order correlation coefficients, multicollinearity was assessed by examining tolerance values for each of the criterion variables. Next, level-1 and level-2 residuals from the models were examined for potential violations of normality, linearity and homogeneity of variance. To examine the normality assumption of level-1 and level-2 residuals, box-and-whisker plots of the residuals were created, and the skewness and kurtosis of the residuals were calculated. Normality, linearity and heteroscedasticity were also examined by plotting the level-1 residuals against the predicted values for reading achievement gain.

Prior to running any exploratory models, it is helpful to clarify the levels of the model as well as the variables at each level. In this study, level-1 refers to variables pertaining to the students while level-2 refers to variables pertaining to the teacher.

Variables for both levels are defined below.

#### Level-1 (Student)

- Student reading achievement
- Student SES
- Student grade
- Student ethnicity

#### Level-2 (Teacher)

- Teacher efficacy for student engagement
- Teacher efficacy for instructional strategies
- Teacher efficacy for classroom management
- Total teacher efficacy

Before discussing each model or family of models in detail, a general overview of the models run is provided in Table 9.

Table 9

*Summary of the Model Structure for each Hierarchical Linear Model*

Model and Purpose	Variables
The Unconditional Model: created to ascertain the ICC which estimates the mean reading achievement gain score across classes; specifies the amount of variability at the student level and at the teacher level; provides a benchmark measure of deviance.	None
Model 1 A-G: Level-1 Predictor Models – depicts the contribution of each level-1 variable without including any level-2 variables; assists in deciding which variables to keep.	Model 1A: Black Model 1B: Hispanic Model 1C: Race <sup>+</sup> Model 1D: Grade <sup>+</sup> Model 1E: SES <sup>+</sup> Model 1F: Gender Model 1G: Race*Grade*SES <sup>+</sup>
Model 2 A-D (run for each IS, CM, SE and TE): Depicts the main effect for each type of efficacy by itself (ie. no level-1 variables) and the two-way interaction of each level-1 variable with each level-2 variable.	Model 2A/IS: IS Model 2B/IS: IS*Race Model 2C/IS: IS*Grade Model 2D/IS: IS*SES This pattern was repeated for each family of models: CM, SE and TE
Model 3 A – F (run for each IS, CM, SE and TE): Main effect and three-way interaction models for Level-1 and level-2 variables.	Model 3A/IS: IS and Grade and SES Model 3B/IS: IS*Grade*SES Model 3C/IS: IS and Race and Grade Model 3D/IS: IS*Race*Grade Model 3E/IS: IS and Race and SES Model 3F/IS: IS*Race*SES This pattern was repeated for each family of models: CM, SE and TE
Model 3 G – H (run for each IS, CM, SE and TE): Main effect and four-way interaction models for Level-1 and level-2 variables.	Model 3G/IS: IS and Race and Grade and SES Model 3H/IS: IS*Race*Grade*SES This pattern was repeated for each family of models: CM, SE and TE

*Note:* At each alteration to the models, decisions regarding which regression coefficients to keep were based upon the changes in the AIC and BIC fit statistics. <sup>+</sup> = analytic variables retained.

*The Unconditional Model.* The first step in creating models was to create an unconditional, or null, model. The unconditional model is a multilevel model with no level-1 or level-2 predictors. It provides the intraclass correlation coefficient (ICC) which estimates the mean student achievement score for reading across teachers. The ICC indicates whether or not multilevel models incorporating variables at the student and teacher levels will be useful. The unconditional model also indicates if there is a statistically significant variability in these means and specifies the amount of variability at the student level and at the teacher level before any predictor or control variables are entered. Additionally, the unconditional model provides a benchmark measure of deviance, a measure of the degree of misfit of the model. All subsequent models can then be compared to this model.

#### Level-1 Model

$$Y_{ij} = \beta_{0j} + r_{ij}$$

where:

$y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\beta_{0j}$  = the intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  classroom

$r_{ij}$  = the error

## Level-2 Model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

where:

$\beta_{0j}$  = the level-1 intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  classroom

$\gamma_{00}$  = the level-2 intercept term

$u_{0j}$  = the level-2 intercept residuals

Mixed Model:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$u_{0j}$  = the level-2 intercept residuals

$r_{ij}$  = the error

In order to answer the research questions, various but similar families of models were run. The first family of models included those models examining teacher efficacy for instructional strategies and its relation to student achievement in reading. The second, third and fourth families of models examined teacher efficacy for instructional strategies, teacher efficacy for classroom management and overall teacher efficacy, respectively, in relation to student achievement in reading while taking the contextual variables into consideration.

*The First Group of Models (Model 1 A-G).* The next step was to analyze a model with all of the level-1 variables fixed, or with the corresponding variance components of



the slopes set at zero. By doing this, it was easier to ascertain the contribution of each level-1 variable.

Level -1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{pij} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\beta_{0j}$  = the intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  classroom

$\beta_{1j}$  = the slope of the regression equation for predicting variable  $X_{pij}$

$X_{pij}$  = the  $p$  explanatory variables at level-1

$r_{ij}$  = the level-1 residual error

Level-2:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{p0}$$

where:

$\beta_{0j}$  = the level-1 intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  class

$\beta_{1j}$  = the level-1 slope of the regression equation for predicting variable  $p$

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{p0}$  = the level-2 slope

$u_{0j}$  = the level-2 intercept residuals

Mixed-Model:

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + u_{0j} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{p0}$  = the level-2 slope of variable  $X_{pij}$

$X_{pij}$  = the  $p$  explanatory variables at level-1

$u_{0j}$  = the level-2 intercept residuals

$r_{ij}$  = the level-1 residual error

In this group of models, it was possible to assess the contribution of individual level-1 explanatory variables. Additionally, since full maximum likelihood (FML) estimation was used, the difference in the deviance of each model created in this step was compared to the deviance of the unconditional model to ascertain whether or not there had been an improvement to the model.

*The Second Group of Models (Model 2 A-D).* In the next group of models, the main effect for each type of efficacy with no level-1 variables was run. This model was labeled Model 2A for each type of efficacy. Once the main effect for each type of efficacy was determined, the two-way interaction of each level-1 variable with each level-2 variable was run.

Level-1:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\beta_{0j}$  = the intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  classroom

$r_{ij}$  = the level-1 residual error

Level-2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

where:

$\beta_{0j}$  = the level-1 intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{01}$  = the coefficient of variable  $W_{qj}$

$W_j$  = the  $q$  explanatory variables at level-2

$u_{0j}$  = the level-2 intercept residuals

Mixed-model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + u_{0j} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{01}$  = the coefficient of variable  $W_{qj}$

$W_j$  = the  $q$  explanatory variables at level-2

$u_{0j}$  = the level-2 intercept residuals

$r_{ij}$  = the level-1 residual error

The following model description is for the two-way interaction of each level-1 variable and each level-2 variable.

Level-1:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\beta_{0j}$  = the intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  classroom

$r_{ij}$  = the level-1 residual error

Level-2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \gamma_{0q} X_{pij} W_j + u_{0j}$$

where:

$\beta_{0j}$  = the level-1 intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{01}$  = the coefficient of variable  $W_{qj}$

$W_j$  = the  $q$  explanatory variables at level-2

$\gamma_{0q}$  = the coefficient of the interaction between  $X_{pij}$  and  $W_j$

$X_{pij}$  = the  $p$  explanatory variables at level-1

$u_{0j}$  = the level-2 intercept residuals

Mixed-model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{0q}X_{pij}W_j + u_{0j} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{01}$  = the coefficient of variable  $W_{qj}$

$W_j$  = the  $q$  explanatory variables at level-2

$\gamma_{0q}$  = the coefficient of the interaction between  $X_{pij}$  and  $W_j$

$X_{pij}$  = the  $p$  explanatory variables at level-1

$u_{0j}$  = the level-2 intercept residuals

$r_{ij}$  = the level-1 residual error

*The Third Group of Models (Model 3 A-F).* The next group of models analyzed the main effects and three-way interactions for the various level-1 and level-2 variables. Table 9 provides a full description of each model. Below is a more general description.

Level-1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{p1ij} + \beta_{2j}X_{p2ij} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\beta_{0j}$  = the intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  classroom

$\beta_{1j}$  = the slope of the regression equation for predicting variable  $X_{p1ij}$

$X_{p1ij}$  = the  $p1$  explanatory variables at level-1

$\beta_{2j}$  = the slope of the regression equation for predicting variable  $X_{p2ij}$

$X_{p2ij}$  = the  $p2$  explanatory variables at level-1

$r_{ij}$  = the level-1 residual error

Level-2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

where:

$\beta_{0j}$  = the level-1 intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  class

$\beta_{1j}$  = the level-1 slope of the regression equation for predicting variable  $p1$

$\beta_{2j}$  = the level-1 slope of the regression equation for predicting variable  $p2$

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{01}$  = the coefficient of variable  $W_j$

$W_j$  = the  $q$  explanatory variables at level-2

$\gamma_{10}$  = the level-2 slope for variable  $p1$

$\gamma_{20}$  = the level-2 slope for variable  $p2$

$u_{0j}$  = the level-2 intercept residuals

Main Effect Mixed-model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X_{p1ij} + \gamma_{20}X_{p2ij} + u_{0j} + r_{ij}$$

Interaction Mixed Model:

$$Y_{ij} = \gamma_{00} + \gamma_{11} X_{p1ij}X_{p2ij}W_j + \gamma_{10}X_{p1ij} + \gamma_{20}X_{p2ij} + u_{0j} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{11}$  = the coefficient of interaction among Level-1 variables  $X_{p1ij}$  and  $X_{p2ij}$  and level-2 variable  $W_j$

$W_j$  = the  $q$  explanatory variables at level-2

$\gamma_{10}$  = the level-2 slope for variable  $X_{p1ij}$

$X_{p1ij}$  = the  $p1$  explanatory variables at level-1

$\gamma_{20}$  = the level-2 slope for variable  $X_{p2ij}$

$X_{p2ij}$  = the  $p2$  explanatory variables at level-1

$u_{0j}$  = the level-2 intercept residuals

$r_{ij}$  = the level-1 residual error

*The Fourth Group of Models (Model 3 G - H).* The final group of models analyzed the main effects and four-way interactions for the various level-1 and level-2 variables. Table 9 provides a full description of each model. Below is a more general description.

Level-1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{p1ij} + \beta_{2j}X_{p2ij} + \beta_{3j}X_{p3ij} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\beta_{0j}$  = the intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  classroom

$\beta_{1j}$  = the slope of the regression equation for predicting variable  $X_{p1ij}$

$X_{p1ij}$  = the  $p1$  explanatory variables at level-1

$\beta_{2j}$  = the slope of the regression equation for predicting variable  $X_{p2ij}$

$X_{p2ij}$  = the  $p2$  explanatory variables at level-1

$\beta_{3j}$  = the slope of the regression equation for predicting variable  $X_{p3ij}$

$X_{p3ij}$  = the  $p3$  explanatory variables at level-1

$r_{ij}$  = the level-1 residual error

Level-2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

where:

$\beta_{0j}$  = the level-1 intercept of the regression equation predicting reading achievement at the end of the study in the  $j^{\text{th}}$  class

$\beta_{1j}$  = the level-1 slope of the regression equation for predicting variable  $p1$

$\beta_{2j}$  = the level-1 slope of the regression equation for predicting variable  $p2$

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{01}$  = the coefficient of variable  $W_j$



$W_j$  = the  $q$  explanatory variables at level-2

$\gamma_{10}$  = the level-2 slope for variable  $p1$

$\gamma_{20}$  = the level-2 slope for variable  $p2$

$u_{0j}$  = the level-2 intercept residuals

Main Effect Mixed-model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{p1ij} + \gamma_{20} X_{p2ij} + \beta_{3j} X_{p3ij} + u_{0j} + r_{ij}$$

Interaction Mixed Model:

$$Y_{ij} = \gamma_{00} + \gamma_{11} X_{p1ij} X_{p2ij} X_{p3ij} W_j + \gamma_{10} X_{p1ij} + \gamma_{20} X_{p2ij} + \beta_{3j} X_{p3ij} + u_{0j} + r_{ij}$$

where:

$Y_{ij}$  = the reading achievement of the  $i^{\text{th}}$  child in the  $j^{\text{th}}$  class

$\gamma_{00}$  = the level-2 intercept term

$\gamma_{11}$  = the coefficient of interaction among Level-1 variables  $X_{p1ij}$ ,  $X_{p2ij}$  and  $X_{p3ij}$  and level-2 variable  $W_j$

$W_j$  = the  $q$  explanatory variables at level-2

$\gamma_{10}$  = the level-2 slope for variable  $X_{p1ij}$

$X_{p1ij}$  = the  $p1$  explanatory variables at level-1

$\gamma_{20}$  = the level-2 slope for variable  $X_{p2ij}$

$X_{p2ij}$  = the  $p2$  explanatory variables at level-1

$\beta_{3j}$  = the slope of the regression equation for predicting variable  $X_{p3ij}$

$X_{p3ij}$  = the  $p3$  explanatory variables at level-1

$u_{0j}$  = the level-2 intercept residuals

$r_{ij}$  = the level-1 residual error

### *3.4.5 Interpretation techniques*

HLM provides a variety of statistical tests for hypothesis testing that were all taken into consideration when interpreting results. T-tests and chi-square tests are provided for the fixed effects, and chi-square tests are provided for the level-2 residual variance. In addition to the t-tests and chi-square tests, each model can be compared to the other models to assess model fit using the deviance measure which indicates a lack of fit between the model and the data; the lower the deviance the better the fit. One disadvantage to deviance, however, is that models with more parameters will always have smaller deviance. That is why it is also important to look at other fit indexes such as the Akaike Information Criterion (AIC) and the Schwarz's Bayesian Information Criterion (BIC). Both of these measures incorporate penalties for a greater number of parameters, creating more parsimonious models (Luke, 2004).

### 3.8 Summary

This chapter outlined the purpose for the study as well as the research questions. Descriptions of the research design including the sample, data collection, dependent and independent variables and the instruments used to measure the variables were also provided. Because the OSTES is a relatively new instrument, there was an outline for the confirmatory factor analysis that was conducted to verify construct validity. The contextual variables of prior achievement, SES, grade and ethnicity were also discussed and a justification for taking them into consideration was proposed. Finally, in the data

analysis section, methodological points related to the use of hierarchical linear modeling, the primary method of analysis used in this study, were described and the models to be run were delineated.

## CHAPTER 4

### RESULTS

This chapter will present descriptive information about the teacher and student participants in the study. It will also outline the results of the confirmatory factor analysis. As discussed in Chapter 3, only one confirmatory factor analysis examined the instrument, and that analysis determined the classroom management factor to be weak (Roberts & Henson, 2001). Tschannen-Moran and Woolfolk Hoy (2001) modified the instrument, but no analysis had been conducted on the modified scale. Prior to running the multilevel analysis, data were analyzed for assumptions, and this analysis is included in the chapter. Finally, the four research questions are addressed through the use of hierarchical linear models, and the results are delineated.

#### 4.1 Purpose of the Study

The purpose of this study was to investigate the effects of teacher efficacy regarding instructional strategies, classroom management and student engagement on ninth and tenth grade students' reading achievement. Teachers with high efficacy in these three areas were hypothesized to positively impact students' reading gains. Student contextual variables of prior student achievement, socioeconomic status (SES), ethnicity and grade were controlled for in the study.

## 4.2 Research Questions

The following research questions were addressed:

1. What is the influence of teacher efficacy regarding instructional strategies on student reading achievement gains?
2. What is the influence of teacher efficacy regarding classroom management on student reading achievement gains?
3. What is the influence of teacher efficacy regarding student motivation on student reading achievement gains?
4. What is the influence of total teacher efficacy (teacher efficacy regarding instructional strategies, classroom management and student motivation combined) on student reading achievement gains?

To address these questions, two levels of data were examined: student level and teacher level. The original student-level variables included gender, ethnicity and socioeconomic status as well as a variable of reading achievement which was created by subtracting 2006 FCAT reading developmental scale scores from 2007 FCAT reading developmental scale scores. Creating this variable allowed the researcher to control for student prior achievement. Development scale scores were used because the scale remains constant over time. During the analysis, it was determined that the inclusion of the variable of gender added little to the study while grade level was a significant variable, so it was added as a variable of interest. Teacher variables included efficacy scores for the three sub-categories of the OSTES, instructional strategies, classroom

management and student engagement, as well as an overall efficacy score that combined the three subcategories.

### 4.3 Descriptive Information

#### 4.3.1. *Sample*

*Students.* Information for the students in this study was provided to the researcher by each of the two counties' Office of Accountability and Research without any identifying features such as student names or student identification numbers. Students were selected for inclusion in the study when one of the teachers responding to the efficacy survey identified the course section identification number of one of his or her ninth or tenth grade English classes. The teacher was asked to contemplate the character of this one identified class while completing the survey.

Information for students in those section identification number classes was obtained from the Offices of Accountability and Research and matched to each participating teacher by the course section identification number. The original data provided by the counties included 2,129 students. The ethnic make-up of the original sample was 57 Asians, 488 Blacks, 441 Hispanics, 1133 Whites, and 10 Others (multi-racial or Eastern Indian). Because the population of Asian and multi-racial or Eastern Indian students was so small, these students were eliminated from the sample, leaving a sample size of 2,062. Additionally, FCAT reading developmental scale scores can only fall in the range of 86 to 3008 points; one score fell outside of this range, and it was considered to be a data entry error and eliminated from the study, leaving a total student

sample size of 2,061. Demographic information for the 2,061 students in the sample is provided in the Table 10.

Table 10

<i>Student Demographic Information</i>		
	<i>n</i>	<i>%</i>
Ethnicity		
Black	488	23.68
Hispanic	440	21.35
White	1133	54.97
Gender		
Female	1020	49.49
Male	1041	50.51
Socioeconomic status		
Not on free and/or reduced lunch	1211	58.76
On free and/or reduced lunch	850	41.24
Course type		
English I	681	33.04
English I Honors	442	21.45
English II	613	29.75
English II Honors	325	15.77
Grade		
Ninth	1123	54.49
Tenth	938	45.51

Overall, the students included in this sample were primarily White (55%). In the multivariate analysis, it was discovered to be more parsimonious to combine Black and Hispanic students into a single category of non-White, thereby making the sample more evenly distributed with 55% White and 45% non-White. In addition, the scores of Black and Hispanic students were similar and behaved in parallel ways; both sets of students scored significantly lower than White students. Therefore, it was determined that there would be no unique contribution by separating out Black and Hispanic students. A more thorough discussion of this decision is outlined later in the chapter. There were slightly

more males than females (51% vs. 49%) and more ninth graders than tenth graders (54% vs. 46%). Forty-one percent of the students were reported to have qualified for free and/or reduced lunch. Students included in the sample were enrolled in one of four types of English classes: English I (33%), English I Honors (21%), English II (30%) and English II Honors (16%).

The created variable of student reading achievement was used as the dependent variable in the study. This variable as well as the developmental scale scores for both the 2006 and 2007 FCAT reading tests are described by grade level in Table 11:

Table 11

*Florida Comprehensive Assessment Test Developmental Scale Scores for Ninth (n = 1123) and Tenth (n = 938) Grade Students*

	Mean	SD	Min	Max
<b>Ninth Grade</b>				
2006 FCAT Scores	1844.79	232.47	886.00	2619.00
2007 FCAT Scores	1897.09	271.57	772.00	2943.00
Reading Achievement	52.30	166.70	-957.00	675.00
<b>Tenth Grade</b>				
2006 FCAT Scores	1899.10	275.99	772.00	2943.00
2007 FCAT Scores	1931.69	295.37	844.00	3008.00
Reading Achievement	32.59	187.98	-773.00	866.00

*Note:* Reading Achievement = the difference in FCAT scores from 2006 to 2007.

It is important to note that developmental scale scores mean different things at different grade levels. In FCAT reading, scale scores, ranging from 100 to 500 for each grade level, are converted to developmental scale scores (DSS or vertical scale scores), which place the scores of students on a scale ranging from 86 to 3008. This continuous



scale allows student progress to be tracked from year to year. In order to create developmental scale scores, anchor items are placed on tests given in different years to relate the scores of the current year to the scores of the first year of operational testing. Hence, the developmental scale score is based on linking items, items that appear identically on the tests of the adjacent grade levels, to relate the scores from one grade to those in the grade above and the grade below.

With the scale score from each grade level linked to those above and below it, a single scale is created. The intended use of the DSS is to monitor the progress of individual students over time. By comparing a student's scores in the same FCAT subject for two or more years, it is possible to determine if a student's performance improved, declined or remained constant. Developmental scale scores were used in this study because the scores and the performance level cut points are constant from year to year; those used for Sunshine State Standard scale scores change from year to year. Developmental scale scores cannot be used to compare across grades, however. For instance, a ninth grader's DSS cannot be compared to a tenth grader's because the two are operating on different scales. Because of this, the student variable of grade level was added as a covariate in the research design. Performance level cut points on developmental scale scores for ninth and tenth grade students are shown in Table 12.

Table 12

*Reading Developmental Scale Scores (86 to 3008) for each Achievement Level on the Florida Comprehensive Assessment Test for Ninth and Tenth Grade*

	Level 1	Level 2	Level 3	Level 4	Level 5
Grade 9	772-1771	1772-1971	1972-2145	2146-2297	2298-2943
Grade 10	844-1851	1852-2067	2068-2218	2219-2310	2311-3008

*Teachers.* The overall sample of teachers included 115 participants. Of the 115 teachers, 76 (66%) had bachelor’s degrees, 38 (33%) had master’s degrees and one (.87%) had a doctorate. Only nine teachers (8%) in the sample held National Board Certification. A majority of the teachers (81 or 70%) became teachers through a traditional teacher education program rather than through alternative means to certification. There were 100 (87%) females and 15 (13%) males. Ethnically, the teacher sample was predominantly white. The mean years of teaching experience was 10.33 ( $SD=9.65$ ). The overall sample of teachers was used to conduct the confirmatory factor analysis.

Five teachers were eliminated prior to running the multivariate analysis because the Office of Accountability could not provide the student data for four of the teachers prior to the running of the analysis, and the fifth teacher was eliminated because no record of her could be found. It was determined that this teacher had likely gotten married and changed her name. The mean years of teaching experience for the subsample was 10.12 ( $SD=9.61$ ) with the minimum years experience being one and the maximum years experience being 35.

Table 13

*Teacher Demographic Data for Confirmatory Factor Analysis  
and Hierarchical Linear Models*

	Sample for CFA ( <i>n</i> =115)		Sample for HLM ( <i>n</i> =110)	
	<i>n</i>	%	<i>n</i>	%
Degree achieved				
Bachelor's	76	66.10	73	66.36
Master's	38	33.04	36	32.73
Doctorate	1	0.87	1	0.91
Route to certification				
Traditional	81	70.43	76	69.09
Alternative	34	29.57	34	30.91
National Board Certification				
Yes	9	7.83	9	8.18
No	106	92.17	101	91.82
Gender				
Male	15	13.04	15	13.64
Female	100	86.96	95	86.36
Ethnicity				
Black	9	7.83	8	7.27
Hispanic	6	5.22	6	5.45
White	97	84.35	94	85.45
Other	3	2.61	2	1.82
Age				
21-25	12	10.43	12	10.91
26-30	32	27.83	30	27.27
31-35	10	8.70	10	9.09
36-40	16	13.91	16	14.55
41-45	4	3.48	4	3.64
46-50	13	11.30	12	10.91
50 and older	28	24.35	26	23.64

#### 4.4 Confirmatory Factor Analysis of the Ohio State Teacher Efficacy Scale

Based on prior evidence and theory, as outlined in Chapter 3, a three-factor model was specified for the Ohio State Teacher Efficacy Scale (see Appendix E) in which questions 1, 2, 4, 6, 9, 12, 14 and 22 loaded onto the latent variable of Efficacy in Student Engagement; questions 7, 10, 11, 17, 18, 20, 23 and 24 loaded onto the latent variable of Efficacy in Instructional Strategies; and questions 3, 5, 8, 13, 15, 16, 19 and 21 loaded onto the latent variable of Efficacy in Classroom Management. The indicators were measured on a scale of 1 – 9, with higher scores reflecting higher levels of efficacy. Table G1 in Appendix G illustrates the correlation matrix of the three-factor model. The measurement model contained no double-loadings. The latent factors of Efficacy in Student Engagement, Instructional Strategies and Classroom Management were permitted to be correlated based on prior evidence of a moderate inverse relationship among these dimensions (Tschannen-Moran & Woolfolk Hoy, 2002). Accordingly, the model was overidentified with 249 degrees of freedom. The fact that the model is overidentified, or there are more elements of the input matrix than freely estimated model parameters, means that goodness-of-fit can be used in the evaluation of the acceptability of the solution. This would not be true if the model were just-identified or underidentified.

As noted in Chapter 3, 115 teachers completed the OSTES. All 115 questionnaires had complete data. Mean scores for the three subscales and for overall teacher efficacy are found in Table 14.

Table 14

*Mean Scores for Ohio State Teacher Efficacy Scale (n = 115)*

	Mean	SD	Min	Max
Teacher Efficacy for Student Engagement (TotSE)	51.97	9.17	27.00	72.00
Teacher Efficacy for Classroom Management (TotCM)	59.80	7.76	35.00	72.00
Teacher Efficacy for Instructional Strategies (TotIS)	58.00	7.76	32.00	72.00
Total Teacher Efficacy (TotTE)	169.77	21.56	96.00	216.00

*Note:* Items on the OSTES are rated on a Likert Scale of 1 – 9 with higher scores indicating higher levels of efficacy. TotSE, TotCM and TotIS are subscales based upon eight questions each with scores ranging from 8 – 72. TotTE is a composite score of the three subscales with scores ranging from 24 – 216.

Before conducting the confirmatory factor analysis, the response distributions for the 24 OSTES items were examined for univariate normality using SAS v9.1.3. Twenty items had both skewness and kurtosis values within  $\pm 1$ . Four items indicated a departure from normality: item 7, *How well can you respond to difficult questions from your students?* (skewness = -1.23, kurtosis = 3.83); item 8, *How well can you establish routines to keep activities running smoothly?* (skewness = -1.11, kurtosis = 0.82); item 23, *How well can you implement alternative strategies in your classroom?* (skewness = -0.74, kurtosis = 1.42); and item 24, *How well can you provide appropriate challenges for very capable students?* (skewness = -1.50, kurtosis = 3.70). Because of this non-normal data and because the sample size was not large ( $n = 115$ ), the sample variance-covariance matrix was analyzed using Mplus version 3.0 (Muthen & Muthen, 1998-2004) with a robust ML estimator, MLM. Each of the three factors was scaled by fixing a factor pattern coefficient to 1.00. Error covariances for the observed items were fixed to zero.

The acceptability of the fitted CFA solution was evaluated on the basis of three aspects: (1) overall goodness-of-fit; (2) the presence or absence of localized areas of strain in the solution, determined by assessing the standardized residuals and the modification indices; and (3) the interpretability, size and statistical significance of the model's parameter estimates. Goodness-of-fit was evaluated using standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), comparative fit index (CFI) and the Tucker-Lewis index (TLI). Following the suggestions of Hu and Bentler (1999), acceptable model fit was defined by the following criteria: RMSEA ( $\leq .06$ ), SRMR ( $\leq .08$ ), CFI ( $\geq .95$ ) and TLI ( $\geq .95$ ). Multiple indices were used because they provide different information about model fit. Used together, these fit indices provide a more conservative and reliable evaluation of the solution. The goodness-of-fit indices suggested that the three-factor model fit the data moderately well, RMSEA = .069, SRMR = .067, CFI = .890 and TLI = .878. While three of the indices fall slightly outside of Hu and Bentler's (1999) cut-off points (RMSEA by .009, CFI by .06 and TLI by .072 respectively), overall the fit is adequate.

An analysis of the standardized residuals indicates the number of standard deviations by which the fitted residuals differ from the zero-value residuals that would be associated with a perfectly fitting model. As seen in Table G2 in Appendix G, the standardized residuals ranged from 0.33 to 0.67. Using the guidelines presented by Brown (2006) suggesting that standardized residuals that are greater than or equal to 2.0 should be considered salient for the presence of localized areas of ill fit, it appears that the standardized residuals support a good fit because none of them is greater than or

equal to 2.0. Modification indices reflect an overall approximation of how much the overall goodness-of-fit would improve if the fixed or constrained parameter estimates were freely estimated. According to Jaccard and Wan (1996), modification indices of 4.0 or greater suggest that the overall fit of the model could be significantly improved ( $p < .05$ ) if the fixed or constrained parameters were freely estimated. In the present model, there were no modification indices above 4.0, further indicating good model fit.

Unstandardized and completely standardized parameter estimates from this solution are presented in Table G3. All freely estimated unstandardized parameters were statistically significant ( $p < .001$ ), and the standardized factor loadings were moderate to high and were significantly different from zero ( $p < .05$ ). Standardized factor loadings ranged from .585 to .818 for Student Engagement, .579 to .799 for Instructional Strategies and .576 to .807 for Classroom Management. Factor loading estimates revealed that the indicators were strongly related to their purported latent factors ( $R^2$  values ranged from .34 to .67), further indicating that the questions on the OSTES are reliable indicators of the constructs of efficacy in student engagement, efficacy in instructional strategies and efficacy in classroom management. Moreover, estimates from the three-factor solution indicate a moderate relationship among the three factors in accordance with previous evidence and theory: instructional strategy with student engagement ( $r = .66$ ), classroom management with student engagement ( $r = .53$ ) and classroom management with instructional strategies ( $r = .70$ ).

#### 4.5 Checking Assumptions for the Hierarchical Linear Models

Prior to conducting any analysis of the data, the data were analyzed for assumptions using SAS v9.1.3. Assumptions for HLMs are similar to those of simple regression with the exception that the assumptions must be examined at each level. In the case of this analysis, the following was assumed:

1. The error associated with level-1 is normal and independent with a mean of 0 in each class and equal variances across classes.
2. Whatever student-level predictors of reading achievement gain are excluded in the model and thereby relegated to the error term  $r_{ij}$  are independent of the other student-level predictors.
3. The residual level-2 error,  $u_{0j}$  and  $u_{1j}$ , is assumed multivariate normal, each with a mean of 0 and some variance and covariance among the random elements.
4. Whatever class-level predictors of reading achievement gain are not included in the model and thereby relegated to the error term  $u_{ij}$  are independent of the other class-level predictors.
5. The error at level-1,  $r_{ij}$ , is independent of the residual class effects.
6. Whatever student-level predictors of reading achievement gain are excluded in the level-1 model and thereby relegated to the error term  $r_{ij}$  are independent of the level-2 predictors in the model. Also, whatever class-level predictors of reading achievement gain are not included in the model and thereby relegated to the level-2 random effects,  $u_{ij}$ , are uncorrelated with the student-level predictors.



Assumptions 2, 4 and 6 focus on the relationship between the variables included in the models and those that are excluded and related to the error terms  $r_{ij}$  and  $u_{ij}$ . Their tenability affects bias in estimations due to model specification. Assumptions 1, 3 and 5 focus only on the random portion of the models  $r_{ij}$  and  $u_{ij}$ , and their tenability affects the consistency of the estimates as well as the accuracy of hypothesis tests and confidence intervals.

In order to test the assumptions, first the data were analyzed for multicollinearity through bivariate examination of independent variables with zero-order correlation coefficients as well as by examination of tolerance values. Bivariate relationships between the variables included in the models are depicted in Table 15.

Table 15

*Bivariate Correlation Matrix for all Criterion and Predictor Variables (n = 2,171)*

	Race	SES	Grade	RdgAch	TotIS	TotCM	TotSE	TotTE
Race	1							
SES	.46	1						
Grade	.03	-.02	1					
RdgAch	-.10	-.08	-.06	1				
TotIS	.01	-.03	-.03	.02	1			
TotCM	.00	-.01	.10	-.00	.68	1		
TotSE	-.12	-.13	-.20	.04	.69	.48	1	
TotTE	-.05	-.07	-.06	.02	.90	.81	.86	1

*Note:* Race: 0 = White, 1 = non-White (Black or Hispanic)

SES: 0 = non-free and/or reduced lunch, 1 = free and/or reduced lunch

Grade: 0 = 9, 1 = 10

RdgAch, TotIS, TotCM, TotSE and TotTE are continuous variables

Only seven bivariate associations had absolute values equal to or greater than .30. Of these seven relationships, none was between the criterion variable of reading achievement and a predictor variable; all were between pairs of predictor variables with the strongest relationships existing predictably among the four measures of teacher efficacy. The relationship between race and SES was also moderate (.46).

Tolerance values were also used to indicate the degree of multicollinearity. They can range from 0 to 1 with values near 1 indicating the low levels of collinearity and high levels of independence. Tolerance values near zero indicate that a variable has little

unique contribution to the models. The tolerance values for the contextual variables at level-1 were as follows: race = .785, SES = .785 and grade = .998. Because these tolerance values ranged from .785 to .998, this indicated a high degree of collinearity and that each variable made a unique contribution to the models. Thus, the analysis of the zero-order correlation coefficients as well as the acceptable tolerance values indicated that there was no evidence of multicollinearity when predicting ninth and tenth grade reading achievement.

Next, level-1 and level-2 residuals were examined for potential violations of normality, linearity and homogeneity of variance. For residuals at both levels, box-and-whisker plots of residuals were analyzed as well as the skewness and kurtosis of the residuals. Examination of the box-and-whisker plots (Appendix H, Figures H1 to H8) and skewness and kurtosis values (Table 16) did not suggest serious violations of normality assumptions.

Table 16

*Distribution of Level-1 and Level-2 Residuals*

	<i>sk</i>	<i>ku</i>
Level-1		
Instructional strategies	-0.18	2.87
Classroom management	-0.17	2.85
Student engagement	-0.18	2.85
Total efficacy	-0.18	2.86
Level-2		
Instructional strategies	-0.01	0.19
Classroom management	-0.00	0.17
Student engagement	0.02	0.15
Total efficacy	0.01	0.17

Level-1 residuals were relatively symmetric and slightly leptokurtic whereas the level-2 residuals were normally distributed. Finally, normality, linearity and heteroscedasticity were also examined by plotting the level-2 residuals against the predicted values for student reading achievement gain (Appendix H, Figures H9 to H12). The resulting scatter plots showed no evidence of heteroscedasticity.

#### 4.6 The Two-Level Hierarchical Linear Model Analysis

The analysis began with the unconditional model, a model with no predictor variables at level one or level two. This model indicated the average class-level reading achievement gain in the entire sample of classes. It also specified the intraclass correlation, or what portion of the total variance occurs between classes. The student level (level-1) covariates of ethnicity, SES, gender and grade were then added to the model to determine the unique contribution of each variable. Then, three different families of models were run for each type of teacher efficacy. The first family of models included the level-2 variable for one type of efficacy to get a baseline model and the ICC, and then the level-1 variables found to be statistically significant in the last group of models, race, SES and grade, were added one at a time to determine individual contribution. The second family of models also included the level-2 variable for one type of efficacy as well as combinations of level-1 variables to test for main effect. The final family of models included the level-2 variable for one type of efficacy and interactions among the level-1 variables. See Table 2 in Chapter 3 for a general overview of the structure of the hierarchical linear models before reading the more detailed descriptions.

*The Unconditional Model.* The unconditional model is meant to be a baseline model that serves two purposes. It informs the researcher of the ICC, and it acts as a benchmark against which other more complex models can be compared. The unconditional model can be viewed as a one-way random effects ANOVA model that expresses the student-level outcome  $Y_{ij}$  using a pair of linked models: one at the student level (level-1) and one at the class level (level-2). At level-1, the student's outcome is expressed as the sum of an intercept for the student's class ( $\beta_{0j}$ ) and the random error ( $r_{ij}$ ) associated with the  $i_{th}$  student in the  $j_{th}$  school:

$$Y_{ij} = \beta_{0j} + r_{ij} \quad \text{where } r_{ij} \sim N(0, \sigma^2)$$

At level-2, the class level intercepts are expressed as the sum of the overall mean ( $\gamma_{00}$ ) and a series of random deviations for that mean ( $\mu_{0j}$ ):

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad \text{where } \mu_{0j} \sim N(0, \tau^2)$$

Substituting the first equation into the second equation yields the multilevel model:

$$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij} \quad \text{where}$$

$$\mu_{0j} \sim N(0, \tau_{00}) \text{ and } r_{ij} \sim N(0, \sigma^2)$$

The model can be split into two parts: a fixed effects part which contains the overall intercept ( $\gamma_{00}$ ) and a random part which contains the two random effects, one for the intercept ( $\mu_{0j}$ ) and one for the within-class residual ( $r_{ij}$ ). The model is fit to data to estimate both the fixed effect,  $\gamma_{00}$ , or the average reading achievement gain in the population, and the two random effects,  $\tau_{00}$  indicating the variability in class means and  $\sigma^2$  indicating the variability in reading achievement gains within classes. Table 17 details summary results from the unconditional model.

Table 17

<i>Fixed and Random Effects for the 2-Level Unconditional Model</i>				
<i>Fixed Effects</i>	Coefficients	SE	t-ratio	<i>p</i>
Intercept, $\gamma_{00}$	44.34	5.11	8.67	<.0001
<i>Random Effects</i>	Variance components	SE	z-ratio	<i>p</i>
Intercept, $\tau_{00}$	1195.29	416.19	2.87	0.0020
Residual, $\sigma^2$	30164	968.81	31.13	<.0001
<i>Model Fit</i>				
AIC	27164.8			
BIC	27170.2			

*Note:* Data represent 110 teachers and 2061 students.

ICC = .034

In the case of this unconditional model, the estimated value of  $\tau_{00} = 1195.29$  and the estimated value of  $\sigma^2 = 30164$ . Hypothesis tests in this section indicate that both variance components are significantly different from zero, suggesting that classes do differ in their average reading achievement gains and that there is even more variation among students within classes since the variance component within classes ( $\sigma^2$ ) is nearly 25 times the size of the variance component between classes ( $\tau_{00}$ ). Since there are no explanatory variables in the unconditional model, the residual variances represent unexplained error variance.

Another way to consider the sources of variation in reading achievement gains is to estimate the intraclass correlation, which indicates what portion of the total variance occurs between classes. Intraclass correlation (ICC) can be determined with this formula:

$$p = \tau_{00} / (\tau_{00} + \sigma^2) = 1195.29 / (1195.25 + 30164) = .038$$

An ICC of .038 indicates that 4% of the reading achievement gain is at the group level; hence, there is a bit of clustering of reading achievement gains within classes, and this suggests that an ordinary least squares analysis of the data would likely yield misleading results. Multilevel analysis is also more appropriate because of the nested nature of the data as well as the fact that students grouped in classes exhibit group dependency. According to Roberts (2007), although a researcher might assume no group dependence when a small ICC is computed for an unconditional model, the degree of dependence is determined by the nature of the variables included in the model. Furthermore, the sample size at level-2 met the *a priori* power requirements, and multi-level analysis allows for the fact that of the 110 classes examined, classes varied in sample size with some classes as small as four students while the largest classes contained 28 students. Finally, HLM provides a better mode of analysis for looking at interactions among variables.

Thus far, only the random effects of the unconditional model have been discussed. It is also important to consider the parameter estimates for the fixed effects. Because this was an unconditional model, there is only one fixed effect which is the intercept with an estimate of 44.34, indicating the average class-level reading achievement gain in the entire sample of classes.

*Adding Student-Level (Level-1) Predictors.* The second group of models contains the student-level (level-1) covariates of ethnicity, socioeconomic status, grade and gender. Summary information for these models can be found in Table 18. The term for

the intercept in Model 1-A, 50.01, estimates  $\gamma_{00}$ , the class mean reading achievement gain when the remaining predictors, or the covariates, are 0.

The regression coefficients for ethnicity were statistically significant. Student ethnicity was broken into three groups: Black, Hispanic and White. In each category, data was dummy coded. For instance, non-Black students were coded as 0, and Black students were coded as 1; non-Hispanic student were coded 0, and Hispanic students were coded as 1; non-White students were coded as 0, and White students were coded as 1. The regression coefficient for Black students is -24.31. This means that on average, Black students' reading achievement gains are 24.31 points lower than those of non-Black students. The standard error of 9.45 yields an observed t-statistic of -2.57 ( $p = .01$ ), which signifies a rejection of the null hypothesis that there is no relationship between a student's ethnicity as Black and reading achievement gains.

Interestingly, the regression coefficient for Hispanic students is -17.74, indicating that, on average, Hispanic students' reading achievement gains are 17.74 points lower than those of non-Hispanic students. The standard error of 9.79 yields an observed t-statistic of -1.81 ( $p = .07$ ), which indicates an acceptance of the null hypothesis that there is no relationship between a student's ethnicity as Hispanic and reading achievement gains. In an effort to create the most powerful and parsimonious models, ethnicity was then redefined as White and non-White, and the new variable of Race was created. White students were coded as 0 and non-White were coded as 1. The regression coefficient for non-White students is -31.83, indicating that, on average, non-White students score 31.83 points lower than White students. The standard error of 8.18 yields an observed t-statistic



of -3.89 ( $p = .0001$ ), indicating that there is indeed a relationship between race and reading achievement gain.

Further support for this new variable leading to more parsimonious models can be found by examining the random effects for Model 1-C which now estimate  $\tau_{00}$  to be 884.35 and  $\sigma^2$  to be 30149. These variance components take on different meanings in this model from the unconditional model because there were no predictor variables in the previous model. Because predictors have been added, these now become conditional components. The component for the variance within classes,  $\sigma^2$ , remains relatively unchanged, only decreasing from 30163 to 30149. The variance component representing variation between classes,  $\tau_{00}$ , on the other hand, has decreased more markedly from 1195.29 to 884.35. The random effects indicate that the covariates explain a fair portion of the class-to-class variation in mean reading achievement gains. Determining how much the variance component for  $\tau_{00}$  has diminished between the two models indicates how much of the variation in class means is explained by the covariates. This can be computed in this manner:

$$(1195.29-884.35)/ 1195.29 = 310.94 / 1195.29 = .26$$

This means that 26% of the explained variation in class mean reading achievement gains is explained by the covariate of race alone. Black and Hispanic were also combined into the single variable of race because it creates a more evenly distributed sample with 54.97% of the sample White and 45.03% of the sample non-White.

Grade was also determined to be a statistically significant variable with a regression coefficient of -19.79. This indicates that tenth graders' reading achievement

gains are on average 19.79 points lower than ninth graders'. Another significant variable was that of SES with a regression coefficient of -22.23, indicating that students receiving free and/or reduced lunch average 22.23 points lower in reading achievement gain. The variable of gender was not statistically significant, so it was not included in future models. This analysis of the level-1 predictors indicates that race (White and non-White), grade and SES are statistically significant variables that will be included in future models. When all three of these variables were included in a model with only level-1 predictors (Model 1-G), an analysis of the variation between classes,  $\tau_{00}$ , indicated that they accounted for 37% of the variation in class means. Hence, this model explains 33% more of the variance than was explained by the unconditional model (ICC of Model G-1 = .37; ICC of the unconditional model = .038 [.37 - .038 = .332]):

$$(1195.29-757.04)/ 1195.29 = 438.25 / 1195.29 = .37$$

Note that SES lost its significance when combined in a model with race and grade. This may have occurred because SES is moderately correlated with race (.46). Because it was statistically significant alone, however, it will remain as a variable in future models. The model fit statistics have also improved from the unconditional model to Model 1-G. The AIC has decreased from 27164.8 to 27126.3, and the BIC has decreased from 27170.2 to 27131.7. In order to start answering the research questions of this study, however, it is necessary to begin adding the level-2 variables for teacher efficacy to the models.

Table 18

*Hierarchical Linear Models with Level-1 Variables Added*

	Model 1A	Model 1B	Model 1C	Model 1D	Model 1E	Model 1F	Model 1G
<i>Fixed Effects</i>							
Intercept	50.01* (5.49)	48.01* (5.43)	58.47* (6.06)	231.21* (95.76)	53.42* (5.97)	38.02* (6.49)	239.89* (89.28)
Black	-24.31* (9.45)						
Hispanic		-17.74 (9.79)					
Race			-31.83* (8.18)				-26.73* (9.0)
Grade				-19.79* (10.12)			-18.93* (9.43)
SES					-22.23* (8.20)		-11.89 (9.00)
Gender						12.27 (7.78)	
<i>Random Effects</i>							
Intercept, $\tau_{00}$	1076.25* (399.16)	1094.99* (406.83)	884.35* (376.77)	1110.98* (405.96)	979.93* (392.54)	1228.47* (420.88)	757.04* (361.66)
Residual, $\sigma^2$	30150* (968.63)	30188* (970.54)	30149* (969.52)	30171* (969.03)	30200* (971.34)	30123* (967.74)	30180* (970.99)
<i>Model Fit</i>							
AIC	27152.0	27155.2	27144.3	27154.6	27151.8	27156.4	27126.3
BIC	27157.4	27160.6	27149.7	27160.0	27157.2	27161.8	27131.7

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. The variable of Race in this table indicates non-White students.

\*Statistically significant – variance estimate and intercept,  $p < .05$ .

To determine the influence of each type of teacher efficacy on student reading achievement gains, 12 multilevel models were examined for each type of teacher efficacy. Tables I1 through I12 in Appendix I contain summary information for all of the models for instructional strategies, classroom management, student engagement and total

efficacy. The intraclass correlations for the baseline models (Models 2-A/IS, 2-A/CM, 2-A/SE and 2-A/TE), the models with only the teacher-level variable for efficacy, for each family of models was relatively small: instructional strategies = .039, classroom management = .039, student engagement = .037 and total efficacy = .038. Using the results of the model building, each of the four research questions is answered. Rather than discussing all 12 multilevel models run for each type of efficacy, the research questions will be answered by focusing on the model of best fit for each type of efficacy. These models are included in the text; the other models can be found in Appendix I.

#### *4.6.1. Research Question 1.*

*What is the influence of teacher efficacy regarding instructional strategies on student reading achievement gains?*

Based upon the results from the model of best fit for the instructional strategies HLMs, as seen in Table 19, the main effect of teacher efficacy for instructional strategies (0.42) was not statistically significantly related to student reading achievement gains ( $p = .5007$ ). Class means are predicted to be less than one point higher for students placed in classes of teachers with high efficacy for instructional strategies. Only the main effects for race and grade were significant. On average, non-White students score 34.60 points lower in reading achievement gains than White students ( $p = .0031$ ), and tenth graders score on average 18.89 points lower than ninth graders in reading achievement gains ( $p = .0464$ ).

Table 19

*Hierarchical Linear Model of Best Fit for Instructional Strategies*

<i>Fixed Effects</i>	Coefficients	SE	t-ratio	<i>p</i>
Intercept, $\gamma_{00}$	216.84*	98.49	2.20	.0298
TotIS	0.42	0.63	0.67	.5007
Race	-34.60*	11.68	-2.96	.0031
Grade	-18.89*	9.48	-1.99	.0464
SES	-21.31	12.96	-1.56	.1001
<i>Fixed Effects-Interaction</i>				
Race*Grade*SES	1.94	1.86	1.04	.2988
<i>Random Effects</i>				
	Variance components	SE	z-ration	<i>p</i>
Intercept, $\tau_{00}$	772.79*	366.21	2.11	.0174
Residual, $\sigma^2$	30176*	971.13	31.07	<.0001
<i>Model Fit</i>				
AIC	27120.7			
BIC	27126.1			

*Note:* Values based on SAS Proc Mixed. Note: Data represent 110 teachers and 2061 students. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

#### *4.6.2. Research Question 2.*

*What is the influence of teacher efficacy regarding classroom management on student reading achievement gains?*

Based upon the results from the model of best fit for the classroom management HLMs, as seen in Table 20, the main effect of teacher efficacy for classroom management (0.06) was not statistically significantly related to student reading achievement gains ( $p = .9239$ ). Class means are predicted to be not even one point higher for students placed in classes of teachers with high efficacy for classroom management. Only the main effects for race and grade were significant. On average, non-White students score 34.45 points lower in reading achievement gains than White students ( $p = .0032$ ), and tenth graders score on average 19.27 points lower than ninth graders in reading achievement gains ( $p = .0435$ ).

Table 20

*Hierarchical Linear Model of Best Fit for Classroom Management*

<i>Fixed Effects</i>	Coefficients	SE	t-ratio	<i>p</i>
Intercept, $\gamma_{00}$	241.44*	94.55	2.55	.0121
TotCM	0.06	0.62	0.10	.9239
Race	-34.45*	11.67	-2.95	.0032
Grade	-19.27*	9.54	-2.02	.0435
SES	-21.54	12.96	-1.66	.0966
<i>Fixed Effects-Interaction</i>				
Race*Grade*SES	1.95	1.87	1.04	.2967
<i>Random Effects</i>	Variance components	SE	z-ration	<i>p</i>
Intercept, $\tau_{00}$	785.34*	367.72	2.14	.0164
Residual, $\sigma^2$	30174*	971.01	31.08	<.0001
<i>Model Fit</i>				
AIC	27121.2			
BIC	27126.6			

*Note:* Values based on SAS Proc Mixed. Note: Data represent 110 teachers and 2061 students. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

#### 4.6.3. Research Question 3.

*What is the influence of teacher efficacy regarding student engagement on student reading achievement gains?*

Based upon the results from the model of best fit for the student engagement HLMs, as seen in Table 21, the main effect of teacher efficacy for student engagement (0.31) was not statistically significantly related to student reading achievement gains ( $p = .5599$ ). Class means are predicted to be less than one point higher for students placed in classes of teachers with high efficacy for classroom management. Only the main effects for race were significant. On average, non-White students score 34.20 points lower in reading achievement gains than White students ( $p = .0035$ ).



Table 21

*Hierarchical Linear Model of Best Fit for Student Engagement*

<i>Fixed Effects</i>	Coefficients	SE	t-ratio	<i>p</i>
Intercept, $\gamma_{00}$	217.17*	101.09	2.15	.0339
TotSE	0.31	0.53	0.58	.5599
Race	-34.20*	11.69	-2.93	.0035
Grade	-18.04	9.68	-1.86	.0625
SES	-21.11	12.98	-1.63	.1039
<i>Fixed Effects-Interaction</i>				
Race*Grade*SES	1.95	1.86	1.04	.2966
<i>Random Effects</i>	Variance components	SE	z-ratio	<i>p</i>
Intercept, $\tau_{00}$	779.22*	366.72	2.12	.0168
Residual, $\sigma^2$	30173*	971.96	31.08	<.0001
<i>Model Fit</i>				
AIC	27121.2			
BIC	27126.6			

*Note:* Values based on SAS Proc Mixed. Note: Data represent 110 teachers and 2061 students. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

#### 4.6.4. Research Question 4.

*What is the influence of total teacher efficacy (efficacy regarding instructional strategies, classroom management and student engagement combined) on student reading achievement gains?*

Based upon the results from the model of best fit for the total efficacy HLMs, as seen in Table 22, the main effect of total teacher efficacy (0.12) was not statistically significantly related to student reading achievement gains ( $p = .5999$ ). Class means are predicted to be not even one point higher for students placed in classes of teachers with high total efficacy. Only the main effects for race and grade were significant. On average, non-White students score 34.45 points lower in reading achievement gains than White students ( $p = .0033$ ), and tenth graders score on average 18.83 points lower than ninth graders in reading achievement gains ( $p = .0477$ ).

Table 22

*Hierarchical Linear Model of Best Fit for Total Efficacy*

<i>Fixed Effects</i>	Coefficients	SE	t-ratio	<i>p</i>
Intercept, $\gamma_{00}$	220.64*	100.42	2.20	.0301
TotTE	0.12	0.23	0.52	.5999
Race	-34.45*	11.69	-2.94	.0033
Grade	-18.83*	9.50	-1.98	.0477
SES	-21.24	12.97	-1.64	.1016
<i>Fixed Effects-Interaction</i>				
Race*Grade*SES	1.93	1.86	1.04	.3008
<i>Random Effects</i>	Variance components	SE	z-ration	<i>p</i>
Intercept, $\tau_{00}$	777.83*	366.88	2.12	.0170
Residual, $\sigma^2$	30175*	971.09	31.07	<.0001
<i>Model Fit</i>				
AIC	27122.9			
BIC	27128.3			

*Note:* Values based on SAS Proc Mixed. Note: Data represent 110 teachers and 2061 students. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

#### 4.6 Summary of Findings

The confirmatory factor analysis of the OSTES indicated that the questions on the measure are trustworthy indicators of the constructs of teacher efficacy in student engagement, efficacy in instructional strategies and efficacy in classroom management. The variables loaded cleanly on the three factors, and the fit indices suggested a moderately good fit. These constructs, however, yielded no statistically significant relationship to student reading achievement gains. In the analysis of the three subscales as well as the analysis of total efficacy, the only variables that were statistically significantly related to student reading achievement gains were race (White vs. non-White) and grade; in the case of teacher efficacy for student engagement, only race was a significant variable.

## CHAPTER 5

### DISCUSSION

#### 5.1 Background of the Study

The reading scores of America's middle school and high school students have shown no improvement at all between the years of 1971 and 2004 (NCES, 2004) and, according to the latest National Assessment of Educational Progress (NAEP) report, more than 8 million of America's middle and high school students are struggling readers (NCES, 2003). When reading scores are disaggregated by groups, the statistics are even more alarming. Only 15% of low-income eighth graders read at grade level, and the achievement gaps between White and Black students, White and Hispanic students and students who were eligible for free or reduced-price lunch and those who were not have shown no significant changes between 1992 and 2005 (Perie, Grigg, & Donahue, 2005).

The fact that there are adolescents who struggle with reading is an issue that many acknowledge; the fact that little was being done to rectify the situation was largely unchallenged until the late 1990s when numerous professional organizations sounded a clarion call for reform with regard to adolescent literacy. The call was heard, and, finally, adolescent literacy is beginning to receive the support it needs both in research and in funding.

At the same time, the standards movement has highlighted the need to employ and retain highly qualified teachers. The interest in this goal has caused the public and researchers alike to ponder what it is that makes a teacher highly qualified. The potential connection between teacher effectiveness, in its newest incarnation of an effective teacher is a teacher who positively influences student outcomes, and student achievement has refocused attention on research in teacher effectiveness. A paradigm shift has occurred in this field of research as researchers are no longer simply focusing on observable teacher actions and, instead, are turning their research efforts to understanding teacher beliefs because it is becoming apparent that many reform initiatives related to the standards movement will fail if teacher beliefs are ignored. Recently researchers have come to view the direction of causality between teacher actions and student behavior as cyclical or circular, rather than unidirectional, whereby a teacher's actions affect student behavior which affects teacher behavior and ultimately student achievement. This new conception of causality, combined with advances in cognitive psychology and a broadening of research paradigms (Fang, 1995), has prompted researchers to become increasingly curious about the domain of teacher thought processes.

One construct of teacher beliefs that holds much promise is the concept of teacher efficacy, grounded in Albert Bandura's socio-cognitive theory. Bandura (2001) notes that "[a]mong the mechanisms of personal agency, none is more central or pervasive than people's belief in their capacity to exercise some measure of control over environmental events...It is partly on the basis of efficacy beliefs that people choose what challenges to undertake, how much to expend in the endeavor, how long to persevere in the face of

obstacles and failures, and whether failures are motivating or demoralizing” (p. 10). Tschannen-Moran and Woolfolk Hoy (2001) translated this broad concept of efficacy to teachers to define teacher efficacy as a teacher’s “. . . judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who may be difficult or unmotivated” (p.3). Researchers have found a positive relationship between student achievement and three kinds of efficacy beliefs: students’ self-efficacy beliefs (Pajares, 1996), teachers’ self-efficacy beliefs (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) and teachers’ collective efficacy beliefs about school (Goddard, Hoy, & Woolfolk Hoy, 2000).

While many exciting strides are being made to improve adolescent literacy, we are currently at a crossroads in the journey. One path leads to “routiniz[ing] teaching to the extent that teachers could be taught recipe-like pedagogical methods, adherence to which could be closely monitored and regulated” (Nespor, 1987, p. 326). Many districts are choosing this path to address adolescent literacy concerns by implementing district-wide adopted curricula that are highly prescriptive. The teacher-proof approach, however, minimizes teacher-level decisions about reading and characterizes teachers as technocrats of the state who merely “deliver” a mandated curriculum (Allington, 2002). Holding teachers highly accountable while giving them little autonomy lessens the likelihood that teachers will accept professional responsibility for providing effective instruction (Bean & Harper, 2004).

The other path at the crossroads leads to the study and understanding of teacher beliefs in an effort to improve student achievement through meaningful professional

development. The effectiveness of investing in teachers rather than in teacher-proof curricula was the subject of Darling-Hammond's research (1999) which found that achievement levels in reading have risen in states that invested heavily in teacher development while little progress is evident in states that invested heavily in testing and curriculum standards. The influence of teachers on student learning is a vital component of the discussion of how to improve student achievement in reading.

To date, however, there have only been five studies directly examining the connection between teacher efficacy and student achievement in reading: three of these studies involved elementary-aged children (Tracz & Gibson, 1986; Anderson, Green & Loewen, 1988; Borton, 1991); one examined junior high students (Armor et al., 1979); and one focused on basic skills students in grades 9 and 10 (Ashton and Webb, 1986). The existing research employed single-level correlation analysis, did not take student contextual variables into consideration and utilized measures of efficacy not sensitive to specific tasks of teaching.

## 5.2 Purpose of the Study

The purpose of this study was to investigate the effects of teacher efficacy for instructional strategies, classroom management and student engagement on ninth and tenth grade students' reading achievement. This study investigated students of varying abilities in grades 9 and 10, employed HLM, factored in student contextual variables and utilized a measure of efficacy that is context specific. Teachers with high efficacy in these three areas were hypothesized to positively impact students' reading gains. Student



contextual variables of prior achievement, socioeconomic status (SES), ethnicity and grade were controlled for in the study.

### 5.3 Research Questions

To expand upon existing research in the relationship between teacher efficacy and student achievement in reading, the following research questions were investigated:

1. What is the influence of teacher efficacy regarding instructional strategies on student reading achievement gains?
2. What is the influence of teacher efficacy regarding classroom management on student reading achievement gains?
3. What is the influence of teacher efficacy regarding student motivation on student reading achievement gains?
4. What is the influence of total teacher efficacy (teacher efficacy regarding instructional strategies, classroom management and student motivation combined) on student reading achievement gains?

This chapter will discuss the findings of each question, address the limitations of the research, outline the implications of the findings and make recommendations for future research on the study of teacher efficacy and student achievement in reading.

### 5.4 Discussion of Findings

First, a confirmatory factor analysis was conducted to ascertain the validity of the OSTES. The confirmatory factor analysis indicated that the questions on the measure are dependable indicators of the constructs of teacher efficacy in student engagement, efficacy in instructional strategies and efficacy in classroom management. These

constructs, however, yielded no statistically significant relationship to student reading achievement gains. As a matter of fact, class means on the reading portion of the FCAT for students placed in classrooms with teachers with high efficacy are predicted to be less than one point higher than those of students in classes with teachers of low efficacy. In the analysis of the three subscales as well as the analysis of total efficacy, the only variables that were statistically significantly related to student reading achievement gains were race and grade; in the case of teacher efficacy for student engagement, however, only race was a significant variable.

The findings of this investigation mirror the findings of the studies done by Ashton and Webb (1986) with basic skills students in grades 9 and 11 and Borton (1991) with students in grades 3 and 4 in that no significant relationship was found between teacher efficacy and student reading achievement. In the study conducted by Anderson, Green and Loewen (1988), a significant relationship was found between these two variables for students in grade 3 but not in grade 6. Based upon this study's similar findings, it appears that the influence of teacher efficacy on student achievement in reading may diminish between the elementary school level and the middle and high school level.

The fact that the sample of students in this study was all ninth and tenth grade students could have heavy bearing on the findings for two reasons. First, reading achievement gain is harder to detect as students advance in years because the amount of reading gain becomes incrementally smaller for many students. While some students did show sizeable gains and losses in reading achievement, the average mean reading

achievement gain was quite small overall: for ninth graders in the study sample the mean gain was 52.30 (SD=166.70), and the average mean reading achievement gain for tenth graders in the study sample was 32.59 (SD=187.98). Most students at the high school level, even struggling readers, have mastered the fundamental concepts of reading such as decoding, phonics and reading fluency, for the most part. The types of gains that are made during high school involve comprehension of increasingly difficult texts, critical evaluation of texts and the ability to understand and synthesize multiple points of view (Chall, 1983). Students at the high school level do not advance as rapidly in reading skills as do emergent readers. Furthermore, it is more difficult to detect the influence of teacher efficacy on honors students as most honors students consistently score high on reading achievement tests, and their reading gains as measured by the FCAT are small and more difficult to discern.

Secondly, the influence of teachers upon students decreases as students advance in grade due to the structure of the school day where students in grade 5 and lower spend the entire day with one teacher. The influence of an elementary teacher could be much stronger than that of the middle or high school teacher who only spends 55 minutes or less per day with the students. Even in the middle school environment, students often belong to teams which organize teachers and students into small groups where students travel with the same small group of peers from teacher to teacher. Teachers work closely together to plan interdisciplinary units, to work on scheduling issues and to monitor student progress. These support structures disappear once the student enters high school and teacher contact time is more limited (Wormeli, 2003).

Additionally, great physical, cognitive and emotional changes take place in students beginning in the middle school years and continue into the high school years that result in the diminishing influence of teachers upon students. Among the physical changes that occur are huge spikes in hormone levels (Nottelmann et al., 1987). These changes result in the development of primary and secondary sex characteristics, and students become more concerned with their physical appearance and self-image. Cognitively, students undergo great changes as well as they enter what Piaget defined as the formal operational stage of intellectual development and are able to think in hypothetical ways about abstract thoughts (1977). For instance, they can focus more on the future and tend to become preoccupied with life plans and develop a greater curiosity and concern for the world around them. Adolescents also embark on their search for a personal identity. According to Erikson, adolescents are trying to integrate their childhood experiences with their developing bodies and biological drives, their new thinking capacities and their ever expanding social roles into a comprehensive sense of self (1968). At this time, peer interaction becomes more important, and the role of authority figures such as parents and teachers diminishes (Knowles & Brown, 2000).

### 5.5 Limitations of the Study

The results of this study must be interpreted with caution. Because the study was limited to teachers of and students in regular and honors English Language Arts classes in grades 9 and 10 who attended schools in two different school districts on the west coast of Florida, the finding may not be generalizable to teachers of and students in other grade levels and courses nationwide.

Also, the dependent variable in the study was measured with one single item, the developmental scale scores on the reading portion of the FCAT. It would be preferable to measure reading achievement gain with several instruments that could be tested for internal validity and consistency. Thus, the findings of this study may not generalize to other measures of student achievement in reading.

Additionally, only a portion of the variance in reading achievement gain was accounted for by the variables examined in each model. It is possible that variables of importance are missing from the models. The exclusion of such variables can lead to model misspecification or biased coefficients resulting in potentially misleading findings. The researcher attempted to avoid model misspecification by including the contextual variables of race, grade, SES and by measuring for reading achievement gain from 2006 to 2007 rather than overall reading achievement.

Another limitation of the study is that the construct of teacher efficacy cannot be measured directly. As a result, this study relied on self-report data. Anastasi (1982) suggests that self-report measures can be subject to inaccurate evaluations by respondents. Incidents of this, however, have not been a topic of concern in the literature regarding the measurement of teacher efficacy. Furthermore, on a variable that is internal to the respondent, self-report data are considered more reliable than third-party observations (Starnaman & Miller, 1992).

Finally, efficacy scores for the three subcategories (scores range from 8-72) and total efficacy (scores range from 24-216) showed little variance. The mean scores were as follows: student engagement=51.97 ( $SD = 9.17$ ), classroom management = 59.80 ( $SD =$

7.76), instructional strategies = 58.00 ( $SD = 7.76$ ) and total efficacy = 169.77 ( $SD = 21.56$ ). According to the description on the survey, this means that most teachers rate themselves at about a 6 or a 7 on the 1 to 9 scale where a 1 indicates that a teacher feels that they can do “nothing” in the area, a 3 means they can do “very little,” a 5 means they have “some influence,” a 7 means they feel they can do “quite a bit” and a 9 means they feel they can do a “great deal.” The survey results indicate that most teachers felt fairly efficacious. The lack of variance residing among teacher efficacy scores made it difficult for hierarchical linear models to detect statistical significance.

### 5.6 Implications

The most notable implication of the current study is its contribution to the literature on the relationship between teacher efficacy and student achievement in reading. This study investigated students of varying abilities in grades 9 and 10. Previously, only one study focused on high school aged students, and it was limited to students in basic skills classes only (Ashton & Webb, 1986). Furthermore, unlike the previous five studies that used a correlational analysis with stepwise multiple regression, this study improved on the methodological approach as it employed HLM, allowing the researcher to account for the interdependence of the nested nature of classroom data. Most important, perhaps, is the fact that this study factored in student contextual variables. Research that analyzes teacher effect on student achievement outcomes must take into account the differences in student background. If it does not, it “may ultimately discredit the standards movement” according to Ballou, Sanders, and Wright (2003). Finally, this study utilized a measure of efficacy that was context specific. Teachers were

asked to rate their perceived efficacy in three distinct areas: classroom management, student engagement and instructional strategies. And, rather than asking teachers how efficacious they felt in general, teachers were asked to focus on one specific class of students they were teaching at the time of the survey administration. This aspect of the study recognizes the fact that it is not only subject area but also classroom dynamics that plays an important role in teacher efficacy. No other study has focused so closely on making certain that the instrument used to ascertain efficacy measured it in such a context specific way. Because of these parameters, the findings of this study are likely less biased than previous findings. However, even though the current study contributes to our understanding of the relationship between teacher efficacy and student achievement in reading, the correlational design of the study does not allow causal conclusions to be made and, thus, cannot be used to guide policies or programs related to ninth and tenth grade student achievement in reading. Instead, the strongest implications of this study are best discussed in terms of future research.

## 5.7 Recommendations for Future Research

### *5.7.1 Recommendations for Research in Methodology.*

Ashton and Webb (1986) hypothesized that the relationship between teacher efficacy and student achievement is often indirect as highly efficacious teachers behave in certain ways, and those behaviors result in increased student achievement. In preliminary studies to their own research, Ashton and Webb (1984) found that teachers with a high sense of efficacy have high expectations for all students, establish classroom environments that encourage warm interpersonal relationships and promote strong

academic work. Other research on the behaviors of teachers with high efficacy supports this view of Ashton and Webb (1986). Allinder (1994) found that teachers with high efficacy plan more thoroughly and are more organized. Highly efficacious teachers are willing to try new strategies (Guskey, 1988), persist when teaching becomes difficult (Coladaraci, 1992) and work with struggling students longer (Gibson & Dembo, 1984). They are more humanistic in their classroom management style (Woolfolk & Hoy, 1990), evoke trust from their students (DaCosta & Riordan, 1996) and favor student-centered classrooms (Czerniak & Schriver, 1994) as well as activity-based and experiential learning (Enochs, Scharman & Riggs, 1995). These findings suggest that a teacher's sense of efficacy is a mediating factor that contributes to the relationship between teacher behavior and student achievement.

The mediating effect of teacher efficacy as well as the direction of the relationship between teacher efficacy and student achievement are important considerations. Previous theories explaining human sociocognitive functioning have relied heavily on unidirectional causation in which behavior is depicted as being shaped by internal dispositions or environmental influences. In these models, persons and situations affect each other, but their influence on behavior is unidirectional. Bandura's (1977) social cognitive theory, on the other hand, explains sociocognitive functioning in terms of triadic reciprocal causation. In this model, behavior, cognitive and other personal factors, and environmental events all operate as interacting determinants that influence each other bidirectionally. The current study focused solely on the cognitive aspect of efficacy in that it measured only teachers' self-perceptions about their own efficacy. Furthermore, it



provides no insight into the direction of the relationships among teacher behavior, thoughts and environmental influences.

The two previous points, efficacy as a mediating factor between teacher behavior and student achievement and Bandura's (1977) belief in triadic reciprocal causation and the direction of the relationship between efficacy and achievement, indicate that HLM may not be the most telling statistical analysis for this data. Furthermore, future research should include additional data that speak not only to the cognitive aspects of efficacy, teachers' perception of their own efficacy, but also to the other two factors in the triad: teacher behavior and environmental events. With regard to teacher behavior, researchers could gather data similar to the following: how often teachers ask higher order questions, time in the classroom that is student-centered, how often the teacher employs learning activities that are experiential or inquiry driven, teacher management practices that are humanistic and other data that reflect positive teacher behavior. Data on environmental events could include the following: class size, teacher planning time, class scheduling, school morale, perceptions of administrative support, professional development opportunities and other environmental factors that might influence teacher efficacy. The gathering of such data in addition to the self-efficacy survey would paint a more complete picture of the triadic reciprocal causation by illuminating all three parts of the triad: cognition, behavior and environment.

In the analysis of this data, researchers should employ structural equation modeling (SEM) because it would highlight the other variables that influence teacher efficacy and student achievement as well as the direction of the relationships among

those variables. SEM allows the researcher to enter both measured variables and latent variables. Latent variables are hypothetical constructs that cannot be directly measured and are generally made up of numerous measured variables that serve as indicators of the construct. The model then serves as a proposed pattern of directional and nondirectional linear relationship among the measured and latent variables. Directional relationships indicate a directional influence of one variable on another while nondirectional relationships are correlational and imply no direction of influence. SEM has the potential to shed more light on the influence of efficacy on achievement than HLM by allowing the researcher to enter more variables into the modeling process and analyzing the direction of the influence of one variable upon another as well as the correlational relationships among variables.

#### *5.7.2 Recommendation for Research in Policy.*

Recent federal education policies such as the NCLB Act of 2001 have narrowed the definition of “scientifically based research” to research that includes only experimental and semi-experimental designs, hypothesis testing and random assignment. The full definition can be found in Figure 1. The far-reaching effects of the stringent definition of “scientifically based research” (SBR) is apparent in an analysis of the What Works Clearinghouse (WWC) website. Established by the U.S. Department of Education’s Institute of Education Sciences, the WWC is a database of reviews of educational research conducted to improve student outcomes. Eisenhart and Towne (2003) reported that “[a] look at the WWC website...made clear that clinical trials were the model for assessment of effectiveness, and achievement test scores were the outcome

of paramount importance” (34). The design of this study was meant to be in compliance with the parameters established by NCLB.

The term “scientifically based research”:

(A) Means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs; and

(B) Includes research that:

(i) Employs systematic, empirical methods that draw on observation or experiment;

(ii) Involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;

(iii) Relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;

(iv) Is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition;

(v) Ensures that experimental studies are presented in sufficient detail and clarity to allow for replication, or, at a minimum, offer the opportunity to build systematically on their findings; and

(vi) Has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

---

*Figure 1. Definition of scientifically based research in No Child Left Behind Act of 2001 (pp. 126 – 127).*

This definition of scientifically based research was not meant to specify what kind of research would be federally funded, but what type of research would count as evidence to justify federal dollars. In other words, if a researcher wishes to convert meaningful research findings into any type of federally funded programmatic reform effort to improve student learning, the research upon which the program is grounded must be empirically-based, experimental in nature and randomized. It is possible, however, that qualitative methods of inquiry such as case studies, teacher and student interviews, or participatory action research could broaden our understanding of teacher efficacy and the

influence of efficacy on student achievement. Investigating the triadic reciprocal causation that links behavior, cognitive, and other personal factors and environmental events may require researchers to go beyond the positivist epistemology that underlies NCLB's narrow definition of "scientifically based research." It is recommended that more research be conducted to analyze the relationship of teacher efficacy to student achievement and that the research conducted employ mixed methods.

Another item of interest to policy makers is the fact that the teacher efficacy surveys in this study were administered to teachers in both districts three weeks after their students had completed the FCAT. At that time, both counties' administrative offices were working to create merit pay plans. The survey results may have been affected by the timing of the survey administration in conjunction with the completion of standardized testing and the districts' moves toward an accountability system for two reasons.

First, merit pay programs assume a strong relationship between teacher performance and student achievement. The logic presented by proponents of accountability systems is that if a teacher knows what goals to aim for and is equipped with the proper information, he or she will be confident in his or her ability to increase student performance on the standardized test (Mathison & Freeman, 2006). Mathison and Freeman (2006), however, noted that many teachers expressed the belief that standardized testing drives classroom curriculum and that many teachers feel compelled to teach to the test. Webb (2006) conveyed the experience of teachers who reported internal conflicts between the mandated curriculum as demanded by their district's accountability program and their own professional diagnosis of what would best serve

their students' needs. Wiley (2000) reported that as a consequence of this conflict, many teachers:

...are finding that their feelings about themselves, their students, and their profession are more negative over time. These teachers are susceptible to developing chronic feelings of emotional exhaustion and fatigue, negative attitudes toward their students, and feelings of diminishing job accomplishments...(p. 81)

Such powerlessness to implement professional judgment and a lack of autonomy could contribute to stress factors for teachers, impacting teachers' perceived self-efficacy (Grant & Hill, 2006). At the time the survey was administered, the researcher was privy to casual conversations among teachers. It was noted that two particular comments were made by teachers. One teacher stated, "I'm so glad the FCAT is over. I finally feel like I have my classes back and can teach important stuff again." Another teacher was overheard saying, "I almost lost my students about eight days prior to FCAT with all of that drill and kill. I'm so glad to be done with it." It was noted in the limitations sections of this chapter that there was very little variation among teacher efficacy scores. On average, teachers felt quite efficacious. Responses to the survey may have been influenced by the fact that these teachers felt that they once again had profession decision-making capabilities restored to them. Scores might have been very different if the surveys had been administered during the time of "drill and kill." More research needs to analyze the effects of standardized testing and mandated curricula on teacher efficacy.

Proponents of accountability programs also see the merit pay that is attached to many accountability systems as a means of rewarding teaching excellence. Schunk

(1984) argued that rewards can increase perceived efficacy when the reward conveys information about teacher competence. Conversely, rewards can decrease sense of efficacy if the individual views the reward as manipulative. Therefore, a teacher's perception of the intent of the reward determines the effect on self-efficacy. Recent literature on this topic supports the view that most teachers do not view test-driven accountability systems as accurate indicators of teacher performance and view them as punitive measures instead. Accountability programs make strong causal inferences about teacher impact on student achievement and do so using only aggregated test scores as indicators of teacher quality (McCarthy & Lambert, 2006). Teachers in a study conducted by Mathison and Freeman (2006) expressed dismay that "the multiple and complex objectives of schooling [were] reduced to one narrow measure of success" (p. 62).

When merit pay was awarded to teachers in one of the districts included in this study, 66% of the 5,000 teachers who received merit pay worked in the district's most affluent schools, while only 3% worked in low-income schools. Sixty-six percent of the teachers also worked in schools that were already A-rated by the State's grading system. Donna Violette, a teacher who did receive merit pay, was asked how she felt about being judged by student performance. She replied, "My children on that given day showed enough growth to get it. If the test had been given a week later, or a week earlier, who knows?" When asked how she felt about being one of the teachers who received merit pay, Ms. Violette stated that she "found it embarrassing. My colleagues who are exceptional, exceptional teachers did not get it and that was embarrassing" (Stein, 2008). Beliefs such as this indicate that teachers do not see merit pay as a measure of teacher

competence. Research needs to be done to ascertain the impact of merit pay on teacher efficacy.

Critics of accountability challenge assumptions about the relationship between teacher performance, student achievement and test scores, citing that test-driven accountability policies increase teacher demoralization (McNeil, 2000), teacher attrition (National Commission on Teaching and America's Future, 2003) and teacher stress over the legitimacy of their professional decisions (Jeffrey, 2002). Future research must delve more deeply into the causal link between teacher performance and student achievement as well as investigate the effects of test-driven accountability programs and merit pay on teacher perceived sense of self-efficacy.

### *5.7.3 Recommendations for Research in Teacher Preparation and Professional Development.*

It is also important to note that the teacher sample for this study was predominantly White (84% White, 16% non-White) while the student sample was more evenly distributed between White and non-White (55% vs. 45%). Additionally, the only variable that was statistically significantly related to student reading achievement gains was race. In both the Ashton and Webb (1986) and the Tracz and Gibson (1986) studies, a significant correlation between teacher efficacy and math achievement was discovered, but no significance was found between teacher efficacy and reading achievement. Tracz and Gibson (1986) hypothesized that "a difference in beliefs regarding the importance of external factors across math and reading achievement may well reflect a commonly held belief by educators that a disadvantaged home environment results in a child being at a

greater disadvantage in reading achievement than in math achievement” (7). This is a significant statement when the correlation between race and SES (.46) is taken into account. This fairly strong correlation indicates that there is a relationship between race and SES; thus, teachers may view Black and Hispanic students through the same deficit lens and assume, too, that they are at a greater disadvantage in reading achievement as students from lower SES backgrounds.

The statement above also brings to light the power of teacher beliefs. What may be occurring in some classrooms is what Nespors (1987) referred to as existential presumptions, when abstract concepts such as student labels transform into stable, concrete entities that seem immutable. A teacher with a financially impoverished student may experience the transformation of that individual student from being labeled as “impoverished” to the concrete expression of “poor reader.” Worse, the teacher may believe that there is no way to overcome that student’s disadvantaged background. Teacher beliefs may play a role in the fact that only 15% of low-income eighth graders read at grade level, and the achievement gaps between White and Black students, White and Hispanic students and those who were not eligible for free or reduced-price lunch and those who were have shown no significant changes between 1992 and 2005 (Perie, Grigg, & Donahue, 2005).

The majority of teachers in this study, as well as the vast majority of teachers nationwide, represent a mainstream White European American population (Gay, 2000). The demographic landscape of the United States, however, is rapidly changing, and the potentially different cultural backgrounds and experiences of teachers and their students



underscore the need for teachers to teach from a framework of multicultural competence. Gay (2004), Ladson-Billings (2000) and Jordan-Irvine (2002, 2003) have documented that successful educators of minority students demand excellence from their students, believe strongly in the teaching profession, get to know their students on a personal level, acknowledge cultural differences without a deficit perspective, use content that is relevant and employ instructional strategies that are cooperative, active and constructivist. Future research in the area of teacher efficacy and its influence on reading achievement would benefit by examining teacher efficacy for cultural competence.

### 5.8 Unanswered Questions

In reviewing the implications of the current study as well as the recommendations for future research, it is apparent that this study has resulted in more questions than answers. One question that is particularly troublesome, however, fits into neither the implications section nor the recommendations for future research section. The teachers in this sample perceive themselves as being quite efficacious. These same teachers, however, are producing students who, on average, perform at achievement level 2 on the FCAT.

A level 2 indicates that students read below grade level and have limited success with the challenging content of the Sunshine State Standards. At achievement level 3, students demonstrate partial success with the challenging content of the Sunshine State Standards, but performance is inconsistent. A student scoring in level 3 answers many of the test questions correctly but is generally less successful with questions that are the

most challenging (Florida Department of Education, *Reading and Mathematics, Grade 3 - 10; Science, Grades 5, 8 and 11*, 2007).

In the state of Florida in 2007, only 38 % of all students in grades 9 and 10 were performing at or above achievement level 3 on FCAT reading. This percentage has steadily increased since 2001, when only 32 % of Florida's ninth and tenth graders were performing at or above achievement level 3. When this data is broken down by ethnicity, only 50% of White students, 30% of Hispanic students and 18% of African-American students in grades 9 and 10 were performing at or above level 3 as of 2007 (Florida Department of Education, *Reading and Mathematics, Grade 3 - 10; Science, Grades 5, 8 and 11*, 2007).

How can it be that teachers who feel they can do “quite a bit” with regard to classroom management, instructional strategies and student engagement are producing students who are, on average, not even performing at grade level on measures of reading achievement? This overarching question only leads to more questions. Are secondary teachers adequately prepared to assist students in improving their reading skills? Why is there not a stronger relationship between the OSTES and FCAT reading achievement? Is the FCAT a good measure of reading achievement? Does teacher efficacy matter? These questions will be addressed individually.

The first of the subsidiary questions is whether or not secondary language arts teachers are adequately prepared to assist students in improving their reading skills. The answer to this is a resounding “no.” Because the teaching of reading was long considered the responsibility of elementary educators, most pre-service English teacher education

programs focus on methods of teaching writing, literature and grammar, not on the teaching of reading (Irvin, Buehl & Klemp, 2007). Now that educators are more cognizant of the fact that many adolescents experience difficulty when reading, a few colleges have added a reading course to their secondary English education programs. The most widely selected texts for these courses are Kyleene Beers' *When Kids Can't Read What Teachers Can Do: A Guide for Teachers 6-12* and Chris Tovani's *I Read It, But I Don't Get It: Comprehension Strategies for Adolescent Readers*. This is a step in the right direction, but both texts, and the courses that accompany them, limit their focus to reading comprehension which is only one of the five key elements of reading. There is little attention given to phonemic awareness, phonics, fluency or vocabulary, the other four elements of reading. Research by Curtis & Long (1999) found that 1 out of 10 adolescents has serious difficulties identifying words as a result of problems associated with phonological aspects of word analysis. Teachers are still woefully unprepared to meet the needs of these particular students.

Recognizing this as a problem, the state of Florida mandated requirements for teachers of reading to comply with The No Child Left Behind Act. Prior to this mandate, reading courses were often taught by language arts teachers with no specific training in the teaching of reading. The law specified that by July 2006, teachers who taught reading classes must have teacher certification in reading at the secondary level that must include a K-12 reading endorsement, a K-12 reading coverage, or an elementary education certification. The state outlined six programs that enable teachers to obtain the K- 12 reading endorsement. These involve various combinations of completing college course

work, district-offered in-service programs and on-line course work funded by the Department of Education. Each district in the state must have an approved program. Unfortunately, aside from submitting the plan to the Department of Education every five years for approval, there is little supervision of these programs and no research to date on the effectiveness or impact of these programs on the reading skills of adolescents.

Another step taken by the state to assist struggling adolescent readers is the placement of a literacy coach in each middle and high school. The literacy coaches do not actually work with struggling readers within their schools; instead, their purpose is to offer assistance to teachers to improve the teachers' abilities to work with struggling readers. This solution has not been a panacea, however, as little has been done to provide time for coaching to occur. As a matter of fact, in many counties, teachers are losing planning time.

It is apparent that more needs to be done to prepare secondary teachers to work with students with the most challenging reading difficulties. There are three ways this goal can be met. Pre-service teacher programs need to better prepare teachers to teach reading to adolescents, and the course work needs to focus on all five key elements of reading, not just comprehension. English education programs may want to consider two-track programs where students can opt to take extra course work to prepare them for the teaching of reading in addition to the teaching of English / Language Arts. In-service teachers need to be provided with opportunities to achieve reading endorsement through programs that are more closely monitored and evaluated. Finally, teachers who achieve endorsement must be offered continuing professional development by reading coaches

who operate in a system that allows them the time and resources necessary to deliver quality in-service programs.

An additional question is why is there not a stronger relationship between the OSTES and FCAT reading achievement? While the OSTES is not a measure of efficacy specifically designed for teachers of reading or the language arts, its three-component sub-scale structure is appropriate for use in the assessment of efficacy of language arts and reading teachers. In *Developing Engaged Readers in School and Home Communities* (1996), Guthrie, McGough, Bennett and Rice described engaged readers as motivated to read, strategic in methods of achieving comprehension, metacognitive in the construction of new understandings and socially interactive in their literacy endeavors. Similarly, Guthrie and Wigfield (2000) proposed that “engaged readers in the classroom or elsewhere coordinate their strategies and knowledge [cognition] within a community of literacy [social] in order to fulfill their personal goals, desires and intentions [motivation].” The three-component sub-scale structure of the OSTES examines teacher efficacy in the same three areas: teacher efficacy for student engagement (motivation), instructional strategies (cognition) and classroom management (social). Hence, while the OSTES is not intentionally designed for English/Language Arts teaching, it does measure teachers’ capacity in three areas specifically related to the skills employed by engaged readers. Therefore, it seems that there would be some degree of relationship between the OSTES and student FCAT reading achievement.

The FCAT, however, does not call upon the same skills identified by Guthrie and Wigfield (2000) as skills employed by engaged readers. As a matter of fact, there is no

community of literacy created when preparing for or taking the FCAT. Students are not permitted to be socially interactive. As far as motivation and the FCAT are concerned, the primary motivation for high school students is the need to pass the FCAT in order to graduate. Aside from that, the FCAT offers very little in the way of providing students with a way to fulfill their personal goals, desires and intentions. The reading passages on the FCAT do little to foster motivation because they are often not relevant to the needs or interests of the average high school student. For instance, on past FCAT tests, there was a passage from John Haines' memoir, *The Stars, the Snow, the Fire: Twenty-Five Years in the Alaska Wilderness*. The piece is highly metaphorical, comparing snow to a book. It is much in the vein of Thoreau's *Walden* – not the most thrilling read for the average ninth or tenth grader. Another passage was entitled “King of Fibers,” a fairly technical and dry abridgement of Jon Thompson's work published in *National Geographic* magazine, that outlines the history, development and uses of cotton. A third reading task included three short passages about research on the sun's eclipse and solar winds that requires students to compare and contrast the passages. Another reading passage was a short fiction piece called “The Tree” by Linda Marasco about a father who wants his family to assist him in moving a tree from the back yard to the front yard. The 705-word-long piece does little to engage the reader. It is too short to develop character, and the plot is one-dimensional. Thus, it appears that the only congruence between the skills required by the FCAT and those utilized by engaged readers are the skills pertaining to the cognitive aspects of reading, not the social and motivational aspects. Students are required to be strategic in methods of achieving comprehension, to be metacognitive in the construction of new

understandings and to coordinate their strategies and knowledge. They are asked to do this, however, without the support of social and motivational impetus. The OSTES examines the skills a teacher would need to facilitate in order to create engaged readers; the FCAT is not seeking to measure reader engagement, but solely reader cognition. A measure of efficacy that would align more closely with the skills a teacher needs to facilitate high test scores would be a measure not of how to create engaged readers, but how to create good test-takers. Creating good test-takers, however, should not be the only, or even the primary, goal of education.

Does the fact that the FCAT assesses only the cognitive aspects of reading mean that it is not a good measure of reading achievement? No, the FCAT is a good measure of reading achievement, but it is only one measure. While the FCAT has proven to be highly correlated with other tests of reading achievement, such as the SAT 9 (Florida Department of Education, 2004), standardized tests offer a narrow view of a student's capabilities. Students are given two hours and forty minutes on one day of testing to prove that they are reading proficiently. If they perform poorly on the test, they are given the exact same type of test the following year. This approach to student assessment goes against commonly held beliefs about assessment that recommend a variety of activities designed to determine the student level of learning. Assessment is not a one-shot deal.

The stated goal of FCAT testing is to ensure that all students are making academic gains and performing at grade level, yet the "retain and re-test loop" is not congruent with educational best practices. Assessment of every individual student is required, and the initial battery is both state-wide and cost effective. However, if a student performs poorly

on the FCAT, there should be an alternative means of assessment. A portfolio review or a diagnostic reading examination is more costly than FCAT testing, but the alternative is to retain a proficient reader simply because he or she does not perform well on a standardized test.

The fact that this study produced no findings of statistical significance may also cause some readers to wonder if the theory of efficacy in itself is flawed or if teacher efficacy is perhaps not as important a variable as purported in the opening chapters of this study. To make such claims would be erroneous. Theory cannot be overturned based upon the findings of one study. Furthermore, as previously explained, the OSTES examines the skills a teacher needs to facilitate in order to create engaged readers; the FCAT is not seeking to measure reader engagement, but solely reader cognition. To state that teacher efficacy does not matter simply because there was no statistically significant relationship between the OSTES and FCAT reading achievement gain implies that how well a teacher prepares his or her students for a standardized test is the only way to measure teacher effectiveness. To believe in this argument, one must also believe then that a teacher's sole responsibility is to prepare students to take tests. This limits teacher responsibility to a very narrow aspect of education and disregards teacher responsibility to stimulate student motivation, social skills and creativity. In examining ways to improve student achievement in reading, it seems that many would like to find a "magic bullet," a single variable that increases student achievement. This study highlights the fact that improving student achievement is a multifaceted issue, and the answer will be equally as complex.



## 5.9 Conclusions

Albert Bandura's (1977, 1997) social cognitive theory posits that people are capable of human agency and that human agency operates in a process called triadic reciprocal causation. The triad stems from three inter-related forces: behavior, environmental influences and internal personal factors resulting from affective, cognitive and biological processes. Collaboratively, these three forces shape what a person believes about themselves as well as the actions he or she will take. Central to Bandura's social cognitive theory is his concept of self-efficacy, or "belief in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Consistent with the general formulation of the concept of self-efficacy, Tschannen-Moran and Woolfolk Hoy (2001) defined teacher efficacy as a teacher's "judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who may be difficult or unmotivated" (p. 283).

Nespor (1987) believed that there are two approaches to dealing with the power of teachers' beliefs. One approach is to reduce the role played by teacher beliefs by routinizing teaching "to the extent that teachers could be taught recipe-like pedagogical methods, adherence to which could be closely monitored and regulated" (p. 326). The second approach is to investigate teacher beliefs more closely to learn how these beliefs "come into being, how they are supported or weakened, [and] how people are converted to them" (p. 326).

Despite the fact that this study found no link between teacher efficacy for instructional strategies, classroom management or student engagement and student

achievement in reading, the findings are an important contribution to the body of literature on teacher effects on student achievement because it was the first to investigate students of varying abilities in grades 9 and 10, employ HLM, factor in student contextual variables and utilize a measure of efficacy that is context specific. However, because of the relatively small magnitude of the relationships found in the current study, the examination of teacher efficacy and the influence of those beliefs on student achievement is an area that requires further analysis. To routinize teaching to “recipe-like pedagogical methods” contradicts 30 years of research (Armour et al., 1979; Woolfolk & Hoy, 1990; Darling-Hammond, 1999; Biancarossa & Snow, 2004) that indicates that it is the autonomous, expert teacher who increases student learning gains. As more researchers use the findings and recommendations from the current study to inform new investigations of the complex relationship between teacher efficacy and student achievement in reading, teacher educators, policymakers, teachers and administrators will be better informed as they continue to work towards improving reading achievement scores and narrowing the achievement gaps in adolescent literacy.

## REFERENCES

- ACT. (2006). *Reading between the lines: What the ACT reveals about college level readiness in reading*. Iowa City, IA: Author.
- Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972-2000*. Washington, D.C.: U.S. Department of Education.
- Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, 149, 1-43.
- Albion, P.R. (1999). *Self-efficacy beliefs as an indicator of teachers' preparedness for teaching with technology*. Retrieved February 7, 2007, from <http://www.usq.edu.au/users/albion/papers/site99/1345.html>
- Alliance for Excellent Education. (2005). *Adolescent literacy policy update*. Washington, D.C.: Author.
- Alliance for Excellent Education. (2006). *Why the crisis in adolescent literacy demands a national response*. Washington, D.C.: Author.
- Allinder, R.M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education*, 17, 86-95.
- Allington, R.L. (2002). *Big brother and the national reading curriculum: How ideology trumped evidence*. Portsmouth, NH: Heinemann.
- Alvermann, D.E., Hinchman, K.A., Moore, D.W., Phelps, S.F., & Waff, D.R. (2006). *Reconceptualizing the literacies in adolescents' lives*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Alvermann, D.E., & Moore, D. (1991). Secondary school reading. In R. Barr, M. Kamil, P. Mosenthal, and P.D. Pearson (Eds.), *Handbook of reading research*. (Vol. II, pp. 1013-46). New York: Longman, Inc.
- Anastasi, A. (1982). *Psychological testing*. New York: Macmillan.

- Anderson, R., Green, M., & Loewen, P. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. *Alberta Journal of Educational Research*, 34(2), 148-165.
- Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonell, L., Pascal, A., Pauly, E., & Zellman, G. (1979). *Analysis of the school preferred reading programs in selected Los Angeles minority schools* (Report No. R-2007-LAUSD). Santa Monica, CA: Rand Corporation. (ERIC Document Reproduction Service No. ED130243).
- Ashton, P., Buhr, D., & Crocker, L. (1984). Teachers' sense of efficacy: A self-or norm referenced construct? *Florida Journal of Educational Research*, 26(1), 29-41.
- Ashton, P.T., Olejnik, S., Crocker, L., & McAuliffe, M. (1982, March). *Measurement problems in the study of teachers' sense of efficacy*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Ashton, P.T., & Webb, R.B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York: Longman.
- Au, K.H. (2000). A multicultural perspective for policies for improving literary achievement: Equity and excellence. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 835-851). Mahway, NJ: Lawrence Erlbaum Associates.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29 (1), 37-65.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1980). Gauging the relationship between self-efficacy judgment and action. *Cognitive Therapy and Research*, 4, 263-268.
- Bandura, A. (1986). *Social foundation of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28, 117-148.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.

- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1-26.
- Banks, J.A. (1991b). Multicultural education: Its effects on students' ethnic and gender role attitudes. In J.P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 459-469). New York: Macmillan.
- Barfield, V., & Burlingame, M. (1974). The pupil control ideology of teachers in selected schools. *The Journal of Experimental Education*, 42(4), 6-11.
- Bean, T.W., & Harper, H.J. (2004). Teacher education and adolescent literacy. In T.L. Jetton & J.A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 392-415). New York: Guilford Press.
- Behrendt, B. (2004, January 19). Sticking with the ABCs of reading. *St. Petersburg Times*, (Citrus Times, p. 1).
- Biancarosa, G., & Snow, C.E. (2004). *Reading next: A vision for action and research in middle and high school literacy*. Washington, D.C.: Alliance for Excellent Education.
- Blintz, W.P. (1997). Exploring reading nightmares of middle and secondary school teachers. *Journal of Adolescent and Adult Literacy*, 41(1), 12-24.
- Boggiano, A.K., & Katz, P. (1991). Maladaptive achievement patterns in students: The role of teachers' controlling strategies. *Journal of Social Issues*, 47(4), 35-41.
- Borchers, C.A., Shroyer, M.G., & Enochs, L.G. (1992). A staff development model to encourage the use of microcomputers in science teaching in rural schools. *School Science and Mathematics*, 92(7), 384-391.
- Borich, G.D. (1992). *Effective teaching methods* (2<sup>nd</sup> ed.). New York: MacMillan Publishing.
- Borton, W.M. (1991). *Empowering teachers and students in a restructuring school: A teacher efficacy interaction model and the effect on reading outcomes*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Reproduction Document Service No. ED 335341).
- Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, 16(2), 239-253.

- Bryk, A.S., & Raudenbush, S.W. (2002). *Hierarchical linear models: Application and data analysis methods*. Thousand Oaks, CA: Sage.
- Buehl, D. (1999). Profiles in learning in high school: A second look at secondary reading. *Wisconsin State Reading Association Journal*, 42(2), 1-10.
- Burley, W.W., Hall, B.W., Villeme, M.G., & Brockmeier, L.L. (1991). *A path analysis of the mediating role of efficacy in first-year teachers' experiences, reactions and plans*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Calderhead, J. (1988). The contribution of field experience to primary student teachers' professional learning. *Research in Education*, 40, 33-49.
- Calderhead, J., & Robson, M. (1991). Images of teaching: Student teachers' early conceptions of classroom practice. *Teaching and Teacher Education*, 7, 1-8.
- Campbell, T.C., Gillaspay, J.A., & Thompson, B. (1995, January). *The factor structure of the Bem Sex-Role Inventory (BSRI): A confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX. (ERIC Reproduction Document Service No. ED 380-491).
- Center for Workforce Preparation. (2002). *Higher-skills, bottom-line results: A chamber guide to improving workplace literacy*. Washington, D.C.: U.S. Chamber of Commerce.
- Chall, J. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.
- Chall, J. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Clark, C.M., & Peterson, P. (1986). Teachers' thought process. In Wittrock, M.C. (Ed.), *Handbook of research on teaching*, (3<sup>rd</sup> ed., pp. 255-296). New York: Macmillan.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Coladaraci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education*, 60, 323-337.
- Cordova, D.I., & Lepper, M.R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization and choice. *Journal of Educational Psychology*, 88, 715-730.

- Creswell, J.W. (1994). *Research design: qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications.
- Cuban, L. (1995). *How teachers taught: Constancy and change in American classroom 1890-1990*. New York: Teachers College Press.
- Curtis, M.E., & Longo, A.M. (1999). *When adolescents can't read: Methods and materials that work*. Cambridge, MA: Brookline Books.
- Czerniak, C.M., & Schriver, M.L. (1994). An examination of pre-service science teachers' beliefs and behaviors as related to self-efficacy. *Journal of Science Teacher Education*, 5(3), 77-86.
- DaCosta, J.L., & Riordan, G. (1996, April). *Teacher efficacy and the capability to trust*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Darling-Hammond, L. (1997). Toward what end? The evaluation of student learning for the improvement of teaching. In J. Millman (Ed.), *Grading teachers, grading schools. Is student achievement a valid evaluation measure?* (pp. 248-263). Thousand Oaks, CA: Corwin.
- Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle, WA: Center for Teaching Policy, University of Washington.
- Darling-Hammond, L., French, J., & Garcia-Lopez, S.P. (2002). *Learning to teach for social justice*. New York: Teachers College Press.
- Deci, E.L., Schwartz, A.J., Sheinman, L., & Ryan, R.M. (1981). An instrument to assess adults' orientation toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology*, 73, 642-650.
- Deemer, S.A., & Minke, K.M. (1999). An investigation of the factor structure of the Teacher Efficacy Scale. *Journal of Educational Research*, 93(1), 3-10.
- Dei, G.J.S. (1999). Knowledge and politics of social change: The implication of anti-racism. *British Journal of Sociology of Education*, 20(3), 395-410.
- DeMesquita, P.B., & Drake, J.C. (1994). Educational reform and the self-efficacy beliefs of teachers implementing nongraded primary school programs. *Teaching and Teacher Education*, 10, 291-302.

- Dunn, A. (2006, February 7). Students' writing outpaces reading. *The Ledger*. Retrieved on October 16, 2006, from <http://www.theledger.com/apps/pbcs.dll/article?AID=/20060207/NEWS/602070378/1004>
- Eisenhart, M., & Towne, L. (2003). Contestation and change in national policy on "scientifically based" education research. *Educational Researcher*, 32(7), 31-38.
- Emmer, E. (1990, April). *A scale for measuring teacher efficacy in classroom management and discipline*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (Revised June, 1990).
- Emmer, E.T., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement*, 51, 755-765.
- Enochs, L.G., Scharmann, L.C., & Riggs, I.M. (1995). The relationship of pupil control to preservice elementary science teacher self-efficacy and outcome expectancy. *Science Education*, 79(1), 63-75.
- Evans, E.D., & Tribble, M. (1986). Perceived teaching problems, self-efficacy and commitment to teaching among preservice teachers. *Journal of Educational Research*, 80(2), 81-85.
- Fang, Z. (1995). On paradigm shift in reading/literacy research, *Reading Psychology*, 16, 2, 215-260.
- Feiman-Nemser, S., McDiarmid, G.W., Melnick, S., & Parker, M. (1988). *Changing beginning teachers' conceptions: A study of introductory teacher education courses*. East Lansing, MI: National Center for Research on Teacher Education, Michigan State University.
- Flesch, R. (1955). *Why Johnny can't read: And what you can do about it*. New York: Harper & Row.
- Florida Department of Education. (1998). *Technical report: Florida Comprehensive Assessment Test (FCAT): 1998*. Unpublished. Tallahassee, FL.: Author.
- Florida Department of Education. (2001). *Technical report: For operational test administrators of the 2000 Florida Comprehensive Assessment Test*. Tallahassee, FL: Harcourt Educational Measurement.
- Florida Department of Education. (2004). *Assessment & accountability briefing book*. Tallahassee, FL: State of Florida.



- Florida Department of Education. (2006). *Understanding FCAT reports 2006*. Tallahassee, FL: State of Florida.
- Florida Department of Education. (2007). *NCLB school restructuring*. Tallahassee, FL: State of Florida.
- Florida Department of Education. (2007). *Reading and mathematics, grade 3 - 10; science, grades 5, 8 and 11*. Tallahassee, FL: State of Florida.
- Flowerday, T., & Schraw, G. (2000). Teacher beliefs about instructional choice: A phenomenological study. *Journal of Educational Psychology, 92*(4), 634-645.
- Forsyth, P.B., & Hoy, W.K. (1978). Isolation and alienation in educational organizations. *Educational Administration Quarterly, 14*, 80-96.
- Freeman, D.J., & Porter, A.C. (1989). Do textbooks dictate the content of mathematics instruction in elementary schools? *American Educational Research Journal, 26*, 403-421.
- Gay, G. (2000). *Culturally responsive teaching: Theory, practice and research*. New York: Teachers College Press.
- Gay, G. (2004). *Becoming multicultural educators: Personal journey toward professional Agency*. San Francisco, CA: Jossey Bass.
- Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*, 569-582.
- Gillaspy, J.A. (1996, January). *A primer on confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 395-040).
- Goddard, R.D., Hoy, W.K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*, 479-508.
- Goodenow, C. (1993b). The psychological sense of school membership among adolescents: scale development and educational correlates. *Psychology in the Schools, 30*, 79-90.
- Goodman, J. (1988). Constructing a practical philosophy of teaching: A study of pre-service teachers' professional perspectives. *Teaching and Teacher Education, 4*, 121-137.

- Gorsuch, R.L. (1983). *Factor Analysis* (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Grant, M., & Hill, J. (2006). Weighing the risks with the rewards: Implementing student-centered pedagogy within high stakes testing. In Lambert, R. & McCarthy, C. (Eds.), *Understanding teacher stress in an age of accountability* (pp. 19-42). Greenwich, CN: Information Age Publishing.
- Grossman, P.L. (1989). *What are we talking about anyway? Subject matter knowledge of secondary language arts teachers*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Gurney, E. (2006). *The educational alliance at Brown University*. Retrieved September 15, 2006, from <http://www.alliance.brown.edu/topics/literacy.shtml>
- Guskey, T. R. (1981). Measurement of the responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education*, 32(3), 44-51.
- Guskey, T.R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, 4, 63-69.
- Guskey, T.R., & Passaro, P.D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal*, 31, 627-643.
- Guthrie, J.T., McGough, K., Bennett, L, & Rice, M.E. (1996). Concept-Oriented Reading Instruction: An integrated curriculum to develop motivations and strategies for reading. In L. Baker, P. Afflerback, & D. Reinking (Eds.), *Developing engaged readers in school and home communities* (pp.165-190). Mahwah, NJ: Lawrence Earlbaum Associates.
- Guthrie, J.T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. Kamil, R. Barr, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. III, pp. 403-425). New York: Longman.
- Hearing at the 106<sup>th</sup> Congress, 1<sup>st</sup> Session. (1990). *Improving student achievement: examining the impact of teacher quality and class size*.
- Henriquez, A. (2005). The evolution of an adolescent literacy program: A foundation's journey. *Reading Research Quarterly*, 40. 376-380.

- Henson, R.K. (2001, January). *Teacher efficacy: Substantive implications and measurement dilemmas*. Paper presented at the annual meeting of the Educational Research Exchange, College Station, TX.
- Henson, R.K., Kogan, L.R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement, 61*(3), 404-420.
- Hock, M.F., & Deshler, D.D. (2003). "No Child" Leaves Behind Teen Reading Proficiency. *The Education Digest, 69*(4), 27-35.
- Hoffman, D.A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*(6), 723-744.
- Hollon, R.E., Anderson, C.W., & Roth, K.J. (1991). Science teachers' conceptions of teaching and learning. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 2, pp. 145-168). Greenwich, CT: JAI Press.
- Hoover-Dempsey, K.V., Bassler, O.C., & Brissie, J.S.(1987). Parent involvement: Contribution of teacher efficacy, school socio-economic status, and other school characteristics. *American Educational Research Journal, 24*, 417-435.
- Houston ties teacher pay to students' scores. (2006, January 13). *The St. Petersburg Times*, p. A13.
- Howard, G. (1999). *We can't teach what we don't know: White teachers, multiracial schools*. New York: Teachers College Press.
- Hox, J. (2002) *Multilevel analysis*. Mahwah, NJ: Lawrence Earlbaum Associates.
- Hoy, W.K., & Woolfolk, A.E. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal, 93*, 356-372.
- Hughes, J., Grossman, P., & Barker, D. (1990). Teachers' expectancies, participation in consultations, and perceptions of consultant helpfulness. *School Psychology Quarterly, 5*(3), 167-179.
- Human Resources Research Organization. (2006). Reading and mathematics: Technical report for 2005 FCAT test administrations. San Antonio, TX: Harcourt.
- Hunt, J.B. (2003). Unrecognized progress. *Education Next, 3*(2), 13-27.
- International Reading Association. (2005). *Standards for literacy coaches*. Newark, DE: Author.

- Irvin, J.L., Buehl, D.R., & Klemp, R.M. (2007). *Reading and the high school student: Strategies to enhance literacy*. New York: Pearson Education, Inc.
- Jackson, P.W. (1968). *Life in classrooms*. New York: Holt, Rhinehart and Winston.
- Jeffrey, B. (2002). Performativity and primary teacher relations. *Journal of Educational Policy*, 17(5), 531-546.
- Jetton, T. (2006). *Adolescent literacy instruction*. Retrieved September 15, 2006, from <http://www.learningpt.org/literacy/adolescent/overview.php>
- Joftus, S. (2002). *Every child a graduate: A framework for an excellent education for all middle and high school students*. Washington, D.C.: Alliance for Excellent Education.
- Jordan, M., Jensen, R., & Greenleaf, C. (2001). "Amidst familial gatherings": Reading apprenticeship in a middle school classroom. *Voices from the Middle*, 8, 15-24.
- Jordan-Irvine, J. (2003). *Educating teachers for diversity: Seeing with a cultural eye*. New York: Teachers College Press.
- Jordan-Irvine, J., (Ed.) (2002). *In search of wholeness: African American teachers and their culturally specific practices*. New York: Palgrave.
- Kaczor, B. (2006, February 11). FCAT-based bonus proposed. *Sarasota Herald-Tribune*, (pp, B1- B2).
- Kamil, M.L. (2003). *Adolescents and literacy: Reading for the 21<sup>st</sup> century*. Washington, D.C.: Alliance for Excellent Education.
- Keegan, L.S. (2003). Help wanted. *Education Next* 3(2), 29-31.
- King, P.M., & Kitchener, K.S. (2002). The reflective judgment model: Twenty years of research on epistemic cognition. In B.K. Hofer & P.R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37-61). Mahwah, NJ: Erlbaum.
- Knowles, T., & Brown, D. F. (2000). *What every middle school teacher should know*. Portsmouth, NH: Heinemann.
- Kos, R. (1991). Persistence of reading disabilities: The voices of four middle school students. *American Educational Research Journal*, 28, 875-95.

- Kreft, I. (April, 1992). *The analysis of small group data: A re-analysis of Webb 1982 with random coefficient model*. Paper presented at the Annual Conference of the American Educational Research Association, San Diego, CA.
- Kupermintz, H. (2002). *Teacher effects as a measure of teacher effectiveness: Construct validity considerations in TVAAS (Tennessee Value-Added Assessment Systems)*. (CSE Tech. Rep. 563). Los Angeles, CA: University of California.
- Ladson-Billings, G., (1999). Preparing teachers for diverse student populations: A critical race theory perspective. *Review of Research in Education*, 24, 211-247.
- Ladson-Billings, G. (2000). Fighting for our lives: Preparing teachers to teach African American students. *Journal of Teacher Education*, 51, 3, 206-214.
- Lent, R.W., Brown, S.D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice and performance. *Journal of Vocational Behavior*, 45, 79-122.
- Linn, R.L. (2001). *The design and evaluation of educational assessments and accountability systems*. (CSE Tech. Rep. 539). Los Angeles, CA: University of California.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Luke, D.A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.
- Marsh, H. W., Kong, C., & Hau, K. (2000). *Longitudinal multilevel models of the big fish little fish pond effect on academic self-concept: Counterbalancing contrast and reflected glory effects in Hong Kong schools*. (ERIC Document Reproduction Services No. ED 442 790).
- Mathison, S. & Freeman, M. (2006). Teacher stress and high stakes testing: How using one measure of academic success leads to multiple teacher stressors. In Lambert, R. & McCarthy, C. (Eds.), *Understanding teacher stress in an age of accountability* (pp. 43-64). Greenwich, CN: Information Age Publishing.
- Martin, R.J., & VanGuten, D.M., (2002). Reflected identities: Applying positionality and multicultural social reconstructionism in teacher education. *Journal of Teacher Education*, 53(1), 44-54.

- McCarthy, C., & Lambert, R. (2006). Helping teachers balance demands and resources in an era of accountability. In Lambert, R. & McCarthy, C. (Eds.), *Understanding teacher stress in an age of accountability* (pp. 215-225). Greenwich, CN: Information Age Publishing.
- McCray, A.D., Vaughn, S., & Neal, L.I. (2001). Not all students learn to read by third grade: Middle school students speak out about their reading disabilities. *The Journal of Special Education*, 35(1), 17-30.
- McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Meijer, C.J.W., & Foster, S.F. (1988). The effect of teacher self-efficacy on referral chance. *Journal of Special Education*, 22, 378-385.
- Meltzer, J. (2002). *Adolescent literacy resources: Linking research and practice*. Providence, RI: Northeast and Islands Regional Laboratory at Brown University.
- Midgley, C., Feldlaufer, H., & Eccles, J. (1989). Change in teacher efficacy and student self- and related-beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, 81, 247-258.
- Minke, K.M., Bear, G.G., Deemer S.A., & Griffen, S.M. (1996). Teacher experiences with inclusive classrooms: Implications for special education reform. *The Journal of Special Education*, 30, 152-186.
- Miserandino, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above average children. *Journal of Educational Psychology*, 88, 203-214.
- Mish et al. (Eds.). (2004). *The Merriam-Webster dictionary*. Springfield, MA: Merriam-Webster.
- Moore, D.W., Bean, T.W., Birdyshaw, D., & Rycik, J.A. (1999). *Adolescent literacy: A position statement*. Newark, DE: International Reading Association.
- Morris-Rothschild, B.K., & Brassard, M.R. (2006). Teachers' conflict management styles: The role of attachment styles and classroom management efficacy. *Journal of School Psychology*, 44, 105-121.
- Murdock, T.B., & Miller, A. (2003). Teachers as sources of middle school students' motivational identity: Variable-centered and person-centered analytic approaches. *The Elementary School Journal*, 103(4), pp.383-399.

- National Center for Educational Statistics. (2003). *Remedial education at degree-granting postsecondary institutions in fall, 2000*. Washington, D.C.: U.S. Department of Education.
- National Center for Educational Statistics. (2004). *Long-term trends: Reading*. Washington, D.C.: U.S. Department of Education.
- National Center of Manufacturers, Andersen, & Center for Workforce Success. (2001). *The skills gap 2001*. Washington, D.C.: National Association of Manufacturers.
- National Commission on Teaching and America's Future. (1997). *Doing what matters most: Investing in quality teaching*. New York: Author.
- National Commission on Teaching and America's Future. (2003). *No dream denied: A pledge to America's children*. Washington, DC: Author.
- National Council of Teachers of English. (2004). *A call to action: What we know about adolescent literacy and ways to support teachers in meeting students' needs*. Urbana, IL: Author.
- National Institute of Child Health and Development. (2000). *Report of the National Reading Panel: Teaching children to read, an evidence-based assessment of the scientific research literature on reading and its implication for reading instruction: Report of the subgroups*. Washington, D.C.: National Institute of Child Health and Human Development.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19, 317-328.
- Nieto, S. (1992). *Affirming diversity*. White Plains, NY: Longman.
- Nisbett, R., and Ross, L.(1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Nottelmann, E., Susman, E., Blue, H., Inoff-Germain, G., Dorn, L., Loriaux, D., Cutler, G., & Chrousos, G. (1987). Gonadal and adrenale correlates of adjustment in early adolescence. In Lerner, R. & Foch, T. (Eds.), *Biological-psychological interacts in early adolescence*. Hillsdale, NJ: Erlbaum.
- Oldfather, P., & Dahl, K. (1994). Toward a social constructivist reconceptualization of intrinsic motivation for literacy learning. *Journal of Reading Behavior*, 26, 139-158.

- Oldfather, P., & McLaughlin, H.L. (1993). Gaining and losing voice: A longitudinal study of students' continuing impulses to learn across elementary and middle school contexts. *Research in Middle Level Education*, 3, 1-25.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66, 533-578.
- Perie, M., Grigg, W., & Donahue, P. (2005). *The nation's report card: Reading 2005* (NCES 2006-451). Washington, D.C.: U.S. Department of Education.
- Perie, M., Moran, R., & Lutkus, A.D. (2005). *NAEP 2004 trends in academic progress: Three decades of student performance in reading and mathematics* (NCES 2005-464). Washington, D.C.: U.S. Department of Education.
- Peterson, P.E. (2003). Ticket to nowhere. *Education Next* 3(2), 39-46.
- Piaget, J. (1977a). *The development of thought: Elaboration of cognitive structures*. New York: Viking.
- Pintrich, P.R., & Schunk, D.H. (1996). *Motivation in education: Theory, research and application*. Englewood Cliffs, NJ: Merrill/Prentice Hall.
- Posner, G.J., Strike, K.A., Hewson, P.W., & Gertzog, W.A. (1982). Accommodations of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66, 211-227.
- RAND Reading Study Group. (2002). *Reading for understanding: Toward a RAND program in reading comprehension*. Santa Monica, CA: RAND.
- Rahim, A. (1990). *Rahim organizational conflict inventory-II*. Palo Alto, CA: Consulting Psychologist's Press.
- Ratekin, N., Simpson, M., Alvermann, E., & Dishner, E. (1998). Why teachers resist content reading instruction. *Journal of Reading*, 41(8), 604-609.
- Raudenbush, S., Bryk, A., Cheong, Y.F., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Raudenbush, S., & Liu, X. (2000). Statistical power and optimal design for multi-site randomized trials. *Psychological Methods*, 5(2), 199-213.
- Rich, Y., Lev, S., & Fisher, S. (1996). Extending the concept and assessment of teacher efficacy. *Educational and Psychological Measurement*, 56(6), 1015-1025.



- Riggs, I.M., & Enochs, L.G. (1990). Toward the development of an elementary teacher's science teaching efficacy beliefs instrument. *Science Education*, 74(6), 625-637.
- Roberts, J.K., & Henson, R.K. (2001, April). *A confirmatory factor analysis of a new measure of teacher efficacy: Ohio State Teacher Efficacy Scale*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Roberts, J. K. (2007, March). *Group dependency in the presence of small intraclass correlation coefficients: An argument in favor of NOT interpreting the ICC*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Roeser, R.W., Midgley, C., & Urdan, T.C. (1996). Perceptions of the school psychological climate and early adolescents' psychological and behavioral functioning in school: The mediating roles of goals and belonging. *Journal of Educational Psychology*, 88, 408-422.
- Rokeach, M.(1968). *Beliefs, attitudes and values: A theory of organization and change*. San Francisco, CA: Josey-Bass.
- Rose, J.S., & Medway, F.J. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research*, 74, 185-190.
- Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80, 1-28.
- Ryan, R., Connell, J., & Grolnick, W. (1992). When achievement is not intrinsically motivated: A theory of internalization and self-regulation in school. In K. Boggiano & T. Pittman (Eds.), *Achievement and motivation: A social developmental perspective*. (pp. 167-188). Cambridge, UK: Cambridge University Press.
- Ryan, R.M., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist*, 55, 68-78.
- Saklofske, D.H., Michayluk, J.O., & Randhawa, B.S. (1988). Teachers' efficacy and teaching behaviors. *Psychological Reports*, 63, 407-414.
- Sanders, W. (1998). Value-added assessment. *School Administrator*, 11(55), 24-27.

- School District of Hillsborough County. (2006). *Healthy meal express: frequently asked questions*. Retrieved December 22, 2006, from <http://www.sdhc.k12.fl.us/SNS/faq.asp#1>
- Schunk, D. (1984). Self-efficacy perspective on student behavior. *Educational Psychologist, 19*(1), 48-58.
- Skinner, E.A., Wellborn, J.G., & Connell, J.P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology, 82*, 22-32.
- Sleeter, C.E. (1991). *Empowerment through multicultural education*. Albany, NY: State University of New York Press.
- Sleeter, C.E. (2001). *Cultural difference and power*. New York: Teachers College Press.
- Snijders, T., & Bosker, R. (Fall, 1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18*(3), 237-59.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Snow, C., & Biancarosa, G. (2003). *Adolescent literacy and the achievement gap: What do we know and where do we go from here?* New York: Carnegie Corporation of New York.
- Snow, C.E., Burns, M.F., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.
- Soodak, L.C., & Podell, D.M. (1993). Teacher efficacy and student problems as factors in special education referral. *The Journal of Special Education, 27*(1), 66-81.
- Spybrook, J., Raudenbush, S., Liu, X. & Congdon, R. (2006). *Optimal Design for Longitudinal and Multilevel Research: Documentation for the "Optimal Design" Software*. Retrieved September 22, 2006, from [http://sitemaker.umich.edu/group\\_based/files/odmanual-20060919v176.doc](http://sitemaker.umich.edu/group_based/files/odmanual-20060919v176.doc)
- Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

- Stapleton, C.D. (1997, January). *Basic concepts and procedures of confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 407-419).
- Starnaman, S. M., & Miller, K. I. (1992). A test of a causal model of communication and burnout in the teaching profession. *Communication Education, 41*, 40-53.
- Stefanou, C.R., Perencevich, K.C., DiCintio, M., & Turner, J.C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Journal of Educational Psychology, 39*(2), 97-110.
- Stein, L. (2008, February 24). Merit pay flows to affluent schools. *St. Petersburg Times*, (pp. 1A & 12A).
- Stevens, J.P. (2002). *Applied multivariate statistics for the social sciences* (4<sup>th</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Measurement, 7*, pp. 201-210.
- Sturtevant, E.G. (2004). *The literacy coach: A key to teaching and learning in secondary schools*. Washington, D.C.: Alliance for Excellent Education.
- Sturtevant, E.G., Boyd, F.B., Brozo, W.G., Hinchman, K.A., Moore, D.W., & Alvermann, D.E., (2006). *Principled practices for adolescent literacy: A framework for instruction and policy*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Tabachnick, B.R., & Zeichner, K. (1984). The impact of student teaching experience on the development of teacher perspectives. *Journal of Teacher Education, 35*(6), 28-42.
- Tobin, T.C. (2006, August 23). Reading program poorly managed. *St. Petersburg Times*, (p. 1A).
- Tracz, S.M., & Gibson, S. (1986). *Effects of efficacy on academic achievement*. Paper presented at the annual meeting of the California Educational Research Association: Marina del Rey, CA. (ERIC Document Reproduction Service No. ED281853).
- Tschannen-Moran, M., & Woolfolk Hoy, A. & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202-248.

- Tschannen-Moran, M., Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.
- United States Department of Education. (2006). *Executive Summary of the No Child Left Behind Act*. Retrieved September 10, 2006 from <http://www.ed.gov/nclb/overview/intro/execsumm.html>
- United States Department of Education. (2006). *Overview: Reading First*. Retrieved September 10, 2006, from <http://www.ed.gov/programs/readingfirst/index.html>
- United States Department of Education. (2006). *Overview: Striving Readers*. Retrieved September 10, 2006, from <http://www.ed.gov/programs/strivingreaders/index.html>
- University of Florida. (2000a). *Prototype analysis of school effects*. Value-Added Research Consortium.
- University of Florida. (2000b). *Measuring gains in student achievement: A feasibility study*. Value-Added Research Consortium.
- Vacca, R.T. (1998). Literacy issues in focus: Let's not marginalize adolescent literacy. *Journal of Adolescent and Adult Literacy, 41*(8), 604-609.
- Vacca, R.T., & Alvermann, D.E. (1998). The crisis in adolescent literacy: Is it real or imagined, *NASSP Bulletin, 82*(600), 4-9.
- Webb, P. (2006). The stress of accountability: Teachers as policy brokers in a poverty school. In Lambert, R. & McCarthy, C. (Eds.), *Understanding teacher stress in an age of accountability* (pp. 1-18). Greenwich, CN: Information Age Publishing.
- Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*, 3-25.
- Weinert, F.E., & Helmke, A. (1995). Learning from wise mother nature or big brother instructor: The wrong choice as seen from an educational perspective. *Educational Psychologist, 30*, 135-142.
- Weinstein, C.S. (1989). Teacher education students' preconceptions of teaching. *Journal of Teacher Education, 40*(2), 53-60.
- Wentzel, K.R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology, 89* 411-417.

- Wentzel, K.R. (1998b). School relationships and motivation in middle school: The role of parents, teachers, and peers. *Journal of Educational Psychology, 90*, 202-209.
- Wentzel, K.R. (2002). Are effective teachers like good parents? Teaching style and student adjustment in early adolescence. *Child Development, 73*, 287-301.
- Wertheim, C., & Leyser, Y. (2002). Efficacy beliefs, background variables and differentiated instruction in Israeli prospective teachers. *The Journal of Educational Research, 96*(1), 54-63.
- Wigfield, A., Eccles, J.S., MacIver, D., Reuman, D.A., & Midgley, C. (1991). Transitions during early adolescents: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*, 552-565.
- Wiley, C. (2000). A synthesis of the research on the causes, effects and reduction strategies of teacher stress. *Journal of Instructional Psychology, 27*(2), 80-87.
- Williams G.C., & Deci, E.L. (1996). Internalization of biopsychosocial values in medical students: A test of self-determination theory. *Journal of Personality and Social Psychology, 70*, 767-779.
- Willower, D.J., Eidell, T.L., & Hoy, W.K. (1967). *The school and pupil control ideology*. (Penn State Studies Monographs No. 24). University Park, MD: Pennsylvania State University.
- Wilson, S.M., & Wineburg, S.S. (1988). Peering at history through different lenses: The role of disciplinary perspectives in teaching history. *Teachers College Record, 89*, 525-539.
- Woolfolk, A.E., & Hoy, W.K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*, 81-91.
- Wormeli, R. (2003). *Day one and beyond: Practical matters for new middle-level teachers*. Portland, ME: Stenhouse Publishers.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 2*, pp. 125-145.
- Zeichner, K.M. (1989). Learning from experience in graduate teacher education. In A. Woolfolk (Ed.), *Research perspectives on the graduate preparation of teachers* (pp.12-29). Englewood Cliffs, NJ: Prentice-Hall.

## APPENDICES

## APPENDIX A

### Researcher Script for Department Chairs' Monthly Meeting

Hello, my name is Anete Vásquez. I am a doctoral candidate at University of South Florida and a former English teacher from Pinellas County, and I would like your help in gathering data for my dissertation. The title of my dissertation is *Teacher Efficacy as a Predictor of Achievement in Adolescent Reading*. Teacher efficacy is a teacher's judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who may be difficult or unmotivated. I am trying to ascertain whether there is a relationship between teacher efficacy and student achievement in reading. Past studies have indicated that there is a relationship, but of the five studies conducted, only one involved teachers and students at the high school level. Additionally, none of the studies took students' prior achievement, socioeconomic status, gender or ethnicity into account. My study will.

Helping me gather my data will take less than twenty minutes of your time. I am asking you to have your English I and II teachers fill out a survey. The survey includes one page of demographic information and, on the reverse side, one page which is the *Teachers' Sense of Efficacy Scale*. Ask your teachers to keep one class of ninth or tenth grade students in mind as they fill out the efficacy survey because efficacy is context specific; a teacher may feel very efficacious teaching his or her 3<sup>rd</sup> period tenth grade English class, but not so efficacious teaching his or her 5<sup>th</sup> period Creative Writing class. For purposes of this questionnaire, it does not matter whether a teacher picks a ninth or tenth grade class in which he or she feels especially efficacious; it is more important that the teacher answer the survey honestly with regard to the class he or she choose to keep in mind.

While teachers are asked to provide their name and the course section identification number of one of their English I or II classes, only I, the researcher, will have access to this information. All identifying information will be coded. None of the information will be reported individually. Teachers will return the survey to you, the department chair, in a sealed envelope with his or her name signed across the seal for protection.

You will be asked to distribute a letter from me inviting your teachers to participate in the study along with an informed consent form and the two page survey. The county also requires that you give a copy of the letter from the county granting me permission to conduct this research to each teacher. Teachers will complete the survey and leave it with you in the sealed envelope. You will then return the surveys to me in the postage paid envelope I have provided to you.

Thank you for letting me speak with you today about my research. I hope to discover some exciting findings about the relationship between teacher efficacy and student achievement in reading at the high school level. Thank you.

## APPENDIX B

### Directions for Administering the Survey

1. Read the script below to your English I and II teachers:

Anete Vásquez, a doctoral candidate at University of South Florida, and a former English teacher in Pinellas County would like your help in gathering data for her dissertation. The title of her dissertation is *Teacher Efficacy as a Predictor of Achievement in Adolescent Reading*. Teacher efficacy is a teacher's judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among students who may be difficult or unmotivated. She is trying to ascertain whether there is a relationship between teacher efficacy and student achievement in reading. Past studies have indicated that there is a relationship, but of the five studies conducted, only one involved teachers and students at the high school level. Additionally, none of the studies took students' prior achievement, socioeconomic status, gender or ethnicity into account. Her study will.

Helping Ms. Vásquez gather her data will take less than ten minutes of your time. You will be asked to fill out a survey. The survey includes one page of demographic information and, on the reverse side, one page which is the *Teachers' Sense of Efficacy Scale*. You are asked to keep one class of your ninth or tenth grade students in mind as you fill out the efficacy survey because efficacy is context specific; you may feel very efficacious teaching your 3<sup>rd</sup> period tenth grade English class, but not so efficacious teaching your 5<sup>th</sup> period Creative Writing class. For purposes of this question, it does not matter whether you pick a ninth or tenth grade class in which you feel especially efficacious; it is more important that you answer the survey honestly with regard to the class you do choose to keep in mind.

While teachers are asked to provide their name and the course section identification number of one of their English I or II classes, only the researcher, Ms. Vásquez, will have access to this information. All identifying information will be coded. None of the information will be reported individually.

I am now going to distribute a letter from Ms. Vásquez inviting you to participate in the study, an informed consent form and the two page survey. I am also going to distribute a copy of the letter from Hillsborough/Pinellas County giving Ms. Vásquez permission to conduct this research. If you choose to participate, and I hope you will, please complete this survey now and leave it with me in the sealed envelope.

2. Please return the surveys to Anete Vásquez by April 14, 2007. Once you have received all completed surveys, place them in the addressed envelope provided, and mail them to Anete Vásquez at the University of South Florida.



## APPENDIX C

### Letter to Teacher Participants

Dear Teacher Participant:

You have been invited to participate in a study about the impact of teacher efficacy on student achievement in reading. This study is being conducted by Anete Vásquez, a doctoral candidate at the University of South Florida, as part of her doctoral dissertation. Her faculty sponsor is Dr. Jane Applegate who can be contacted at the University of South Florida, College of Education, 4202 East Fowler Avenue, EDU 162, Tampa, FL 33620. The telephone number is (813)974-3533.

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

1. complete a questionnaire about your teaching background (years of teaching experience, degree achieved, route to certification, and type of certification, etc.);
2. identify one class of ninth or tenth grade students whom you currently teach English I or English II and provide the course section identification number for that class;
3. complete The Teachers' Sense of Efficacy Scale, a survey which measures teacher attitudes about their influence on students' learning while keeping the class of ninth or tenth grade students identified in mind.

Participation in this study will involve approximately 20 minutes of your time. There are no known risks associated with your participation in this research.

Your participation in this study is strictly voluntary but greatly appreciated; there are no consequences for choosing not to participate or for withdrawing from the study. Although there are no monetary rewards, this research may help us to better understand how a teacher's sense of efficacy affects student achievement in reading.

Confidentiality of your identity will be strictly maintained. The information from this study will only be accessible by the researcher. All teacher information will be coded so that no individual can be identified. Similarly, results from the study will be released in a way that no individual answers can be identified. All information gathered will be kept secure, under physical lock and key, for three years after the research has been completed. After three years, it will be destroyed with a paper shredder and recycled.

For any questions regarding this study or problems with your role as a teacher participant, you may contact Anete Vásquez or Dr. Jane Applegate, at the address in the first paragraph.

Sincerely,  
Anete Vásquez

APPENDIX D

Teacher Personal Information Questionnaire

Directions: Please respond to this survey if you teach English I or II (regular or honors).

**Teacher Name:** \_\_\_\_\_

**Teacher School:** \_\_\_\_\_

**1. Please indicate how many years you have taught in a K-12 setting:** \_\_\_\_\_ years

**2. Please indicate, with an X, the highest degree you have attained:**

Bachelor's \_\_\_\_\_ Master's \_\_\_\_\_ Doctorate \_\_\_\_\_

**3. Please indicate, with an X, the type of teacher preparation you received:**

Traditional \_\_\_\_\_ Alternative \_\_\_\_\_

**4. Please indicate, with an X, if you hold National Board Certification:**

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

**5. Please indicate, with an X, your gender:** Male \_\_\_\_\_ Female \_\_\_\_\_

**6. Please indicate, with an X, to which age group you belong:**

21-25 years old \_\_\_\_\_ 26-30 years old \_\_\_\_\_

31-35 years old \_\_\_\_\_ 36-40 years old \_\_\_\_\_

41-45 years old \_\_\_\_\_ 46-50 years old \_\_\_\_\_

50 and older \_\_\_\_\_

**7. Please indicate, with an X, your ethnicity:**

African-American \_\_\_\_\_ Asian-American \_\_\_\_\_

Hispanic \_\_\_\_\_ White \_\_\_\_\_

Other \_\_\_\_\_ (please indicate): \_\_\_\_\_

**8. Please outline your daily teaching schedule this year in the space provided below.**

**Provide periods and complete class titles. (Example: Period I – English I Honors; Period 2 – English III; Period 3 – Yearbook; Period 4 – Planning; Period 5 – English III; Period 6 – Planning; Period 7 – English II Honors):**

**9. Efficacy is believed to be context specific, so as you respond to the items on The Teachers' Sense of Efficacy Scale (on the *back* of this page), keep one class of *ninth* or *tenth* graders to whom you currently teach English I or II in mind. Circle the period and class title of that particular class in the list above and describe the character of the selected English I or English II class in the space below:**

Appendix E

Teachers' Sense of Efficacy Scale

**Directions:** This questionnaire is designed to assess teacher efficacy in the various tasks teachers are asked to do daily. Please indicate your opinion about each of the statements below by clearly circling one number from 1 (Nothing) to 9 (A Great Deal). Your answers are confidential.

	How much can you do?									
	Nothing	Very Little	Some Influence	Quite a Bit	A Great Deal					
1. How much can you do to get through to the most difficult students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
2. How much can you do to help your students think critically?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
3. How much can you do to control disruptive behavior in the classroom?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
4. How much can you do to motivate students who show low interest in school work?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
5. To what extent can you make your expectations clear about student behavior?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
6. How much can you do to get students to believe they can do well in school work?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
7. How well can you respond to difficult questions from your students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
8. How well can you establish routines to keep activities running smoothly?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
9. How much can you do to help your students value learning?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
10. How much can you gauge student comprehension of what you have taught?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
11. To what extent can you craft good questions for your students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
12. How much can you do to foster student creativity?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
13. How much can you do to get children to follow classroom rules?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
14. How much can you do to improve the understanding of a student who is failing?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
15. How much can you do to calm a student who is disruptive or noisy?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
16. How well can you establish a classroom management system with each group of students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
17. How much can you do to adjust your lessons to the proper level for individual students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
18. How much can you use a variety of assessment strategies?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
19. How well can you keep a few problem students from ruining an entire lesson?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
20. To what extent can you provide an alternative explanation or example when students are confused?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
21. How well can you respond to defiant students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
22. How much can you assist families in helping their children do well in school?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
23. How well can you implement alternative strategies in your classroom?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
24. How well can you provide appropriate challenges for very capable students?	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	

## APPENDIX F

### Informed Consent Form

Space below reserved for IRB Stamp –  
Please leave blank

#### **Informed Consent for an Adult**

Social and Behavioral Sciences  
University of South Florida

#### **Information for People Who Take Part in Research Studies**

---

---

The following information is being presented to help you decide whether or not you want to participate in a research study. Please read this carefully.

**Title of research study:** Teacher Efficacy as a Predictor of Student  
Achievement in Reading

**Person in charge of study:** Anete Vásquez

**Where the study will be done:** All high schools in your school district

#### **Should you take part in this study?**

This form tells you about this research study. You can decide if you want to take part in it. You do not have to take part. Reading this form can help you decide.

#### **Why is this research being done?**

The purpose of this study is to gain a better understanding of how teachers' sense of efficacy affects student achievement in reading. Responses will be kept strictly confidential and individual responses will not be identified.

#### **Plan of Study**

The survey includes two sections. The first is a section asking for demographic information. It also asks participants to identify one class to whom he or she teaches English I or II and to keep that class in mind as he or she responds to the *Teachers' Sense of Efficacy Scale*. The survey should take approximately 20 minutes to complete. Your participation is voluntary. Refusal to participate will not result in any penalty or loss of benefits. You may withdraw your participation at any time.

## APPENDIX F (continued)

### **Benefits of Taking Part**

Your responses will help the researcher understand how teacher efficacy affects student achievement in reading. In the past such data has been used to assist districts in creating targeted professional development and classroom materials for teachers.

### **Payment for Participation**

You will not be paid for taking part in this survey.

### **What are the risks if you take part in this study?**

There are no known risks to those who take part in this study.

### **What will we do to keep your study records private?**

Responses will be kept strictly confidential and individual responses will not be identified. Federal law requires us to keep your study records private:

- The USF Institutional Review Board (IRB)
- The United States Department of Health and Human Services (DHHS)

### **Questions and Contacts**

- If you have any questions about this research study, contact Anete Vásquez at (813)974-2209.
- If you have questions about your rights as a person who is taking part in a research study, you may contact the Division of Research Compliance of the University of South Florida at (813) 974-5638. Consent to Take Part in this Research Study

## APPENDIX G

### Additional Confirmatory Factor Analysis Tables

Table G1.

*Correlation Matrix of the Three-factor Model*

	Q1SE	Q2SE	Q3CM	Q4SE	Q5CM	Q6SE	Q7IS	Q8CM	Q9SE	Q10IS	Q11IS	Q12SE
<i>Q1SE</i>	1.000											
<i>Q2SE</i>	0.612	1.000										
<i>Q3CM</i>	0.362	0.249	1.000									
<i>Q4SE</i>	0.641	0.456	0.347	1.000								
<i>Q5CM</i>	0.249	0.282	0.469	0.216	1.000							
<i>Q5SE</i>	0.552	0.478	0.245	0.656	0.257	1.000						
<i>Q7IS</i>	0.292	0.422	0.321	0.190	0.412	0.286	1.000					
<i>Q8CM</i>	0.216	0.270	0.353	0.217	0.563	0.165	0.501	1.000				
<i>Q9SE</i>	0.383	0.380	0.397	0.536	0.159	0.391	0.319	0.297	1.000			
<i>Q10IS</i>	0.247	0.337	0.214	0.239	0.312	0.275	0.336	0.285	0.359	1.000		
<i>Q11IS</i>	0.227	0.427	0.104	0.254	0.412	0.308	0.552	0.390	0.298	0.613	1.000	
<i>Q12SE</i>	0.458	0.407	0.227	0.556	0.227	0.496	0.262	0.295	0.488	0.270	0.297	1.000

Table G1. (continued)

*Correlation Matrix of the Three-factor Model*

	Q13CM	Q14SE	Q15CM	Q16CM	Q17IS	Q18IS	Q19CM	Q20IS	Q21CM	Q22SE	Q23IS	Q24IS
Q13CM	1.000											
Q14SE	0.288	1.000										
Q15CM	0.621	0.442	1.000									
Q16CM	0.626	0.263	0.595	1.000								
Q17IS	0.355	0.414	0.471	0.437	1.000							
Q18IS	0.331	0.340	0.318	0.291	0.440	1.000						
Q19CM	0.550	0.186	0.584	0.544	0.361	0.314	1.000					
Q20IS	0.400	0.321	0.447	0.508	0.454	0.548	0.438	1.000				
Q21CM	0.545	0.274	0.596	0.537	0.344	0.289	0.540	0.446	1.000			
Q22SE	0.156	0.393	0.260	0.267	0.343	0.312	0.279	0.293	0.214	1.000		
Q23IS	0.413	0.359	0.423	0.419	0.548	0.546	0.451	0.540	0.338	0.461	1.000	
Q24IS	0.360	0.276	0.334	0.389	0.379	0.457	0.411	0.491	0.442	0.320	0.634	1.000

APPENDIX G (continued)

Table G2

*Standardized Residuals of the Confirmatory Factor Analysis*

---

Q1SE	0.567
Q2SE	0.438
Q3CM	0.603
Q4SE	0.669
Q5CM	0.403
Q6SE	0.544
Q7IS	0.359
Q8CM	0.332
Q9SE	0.395
Q10IS	0.335
Q11IS	0.507
Q12SE	0.439
Q13CM	0.617
Q14SE	0.498
Q15CM	0.610
Q16CM	0.651
Q17IS	0.410
Q18IS	0.409
Q19CM	0.516
Q20IS	0.531
Q21CM	0.462
Q22SE	0.342
Q23IS	0.639
Q24IS	0.480

---



APPENDIX G (continued)

Table G3.

*Parameter Estimates from the Three-Factor CFA Models of Efficacy in Student Engagement, Instructional Strategies and Classroom Management*

	Estimates	S.E.	Est./S.E.	Std	StdXY
Student Engagement by					
Q1SE	1.000	0.000	0.000	1.250	0.753
Q2SE	0.749	0.079	9.423	0.936	0.662
Q4SE	1.129	0.110	10.229	1.411	0.818
Q6SE	0.817	0.089	9.146	1.022	0.737
Q9SE	0.794	0.101	7.899	0.992	0.629
Q12SE	0.750	0.099	7.553	0.938	0.663
Q14SE	0.778	0.111	7.027	0.973	0.706
Instructional Strategy by					
Q7IS	1.000	0.000	0.000	0.592	0.599
Q10IS	0.987	0.245	4.033	0.585	0.579
Q11IS	1.397	0.225	5.473	0.828	0.712
Q17IS	1.589	0.359	4.422	0.941	0.640
Q18IS	1.49	0.408	3.656	0.883	0.639
Q20IS	1.396	0.203	6.068	0.827	0.729
Q23IS	1.749	0.273	6.572	1.062	0.799
Q24IS	1.629	0.211	7.705	0.965	0.693
Classroom Management by					
Q3CM	1.000	0.000	0.000	1.086	0.777
Q5CM	0.576	0.076	1.604	0.626	0.635
Q8CM	0.567	0.098	5.787	0.615	0.576
Q13CM	0.914	0.093	9.826	0.993	0.786
Q15CM	1.031	0.086	12.025	1.119	0.781
Q16CM	0.894	0.104	8.628	0.970	0.807
Q19CM	0.962	0.097	9.866	1.044	0.718
Q21CM	0.929	0.110	8.465	1.009	0.680
Instructional Strategies with					
Student Engagement	0.486	0.16	3.047	0.657	0.657
Classroom Management with					
Student Engagement	0.716	0.171	4.177	0.528	0.528
Instructional Strategies	0.449	0.138	3.253	0.698	0.698

APPENDIX G (continued)

Table G3.(continued)

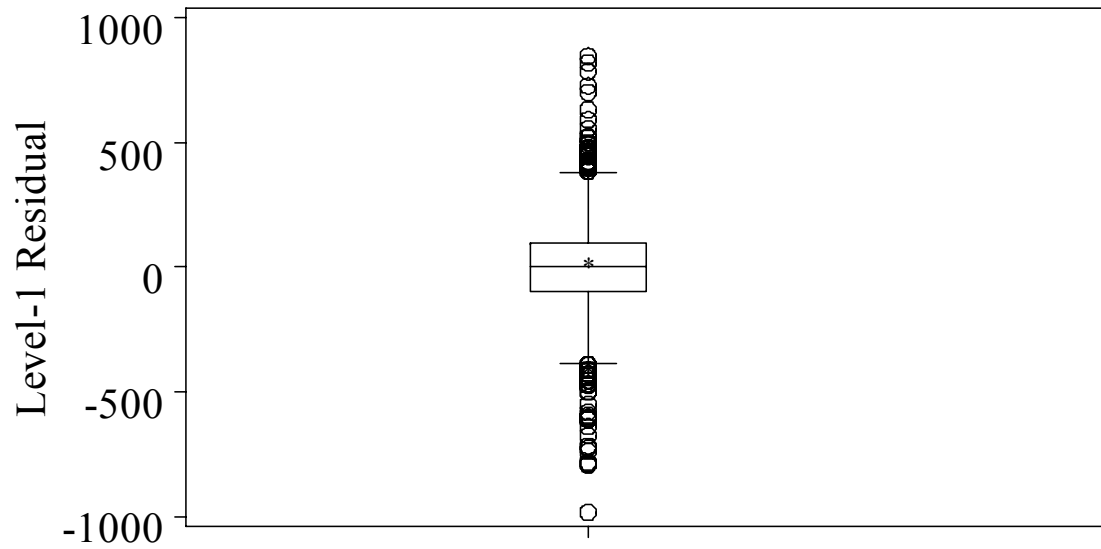
*Parameter Estimates from the Three-Factor CFA Models of Efficacy in Student Engagement, Instructional Strategies and Classroom Management*

	Estimates	S.E.	Est./S.E.	Std	StdXY
Variances					
Student					
Engagement	1.562	0.321	4.864	1.000	1.000
Instructional					
Strategies	0.351	0.158	2.219	1.000	1.000
Classroom					
Management	1.179	0.210	5.605	1.000	1.000
Residual Variances					
Q1SE	1.195	0.125	9.579	1.195	0.433
Q2SE	1.123	0.141	7.950	1.123	0.562
Q3CM	0.775	0.085	9.161	0.775	0.397
Q4SE	0.985	0.180	5.486	0.985	0.331
Q5CM	0.580	0.070	8.239	0.580	0.597
Q6SE	0.876	0.082	10.739	0.878	0.456
Q7IS	0.626	0.061	10.277	0.626	0.641
Q8CM	0.761	0.096	7.921	0.761	0.668
Q9SE	1.508	0.154	9.782	1.508	0.605
Q10IS	0.679	0.061	11.140	0.679	0.665
Q11IS	0.666	0.070	9.531	0.666	0.493
Q12SE	1.122	0.168	6.693	1.122	0.561
Q13CM	0.611	0.078	7.855	0.611	0.383
Q14SE	0.953	0.121	7.882	0.953	0.502
Q15CM	0.800	0.086	9.328	0.800	0.390
Q16CM	0.505	0.061	8.220	0.505	0.349
Q17IS	1.276	0.135	9.455	1.276	0.590
Q18IS	1.127	0.209	5.396	1.127	0.591
Q19CM	1.021	0.164	6.221	1.021	0.484
Q20IS	0.604	0.071	8.557	0.604	0.469
Q21CM	1.183	0.143	8.273	1.183	0.538
Q22SE	2.122	0.237	8.959	2.122	0.658
Q23IS	0.638	0.078	8.157	0.638	0.361

APPENDIX H

Assumption Checking for the HLMs

---

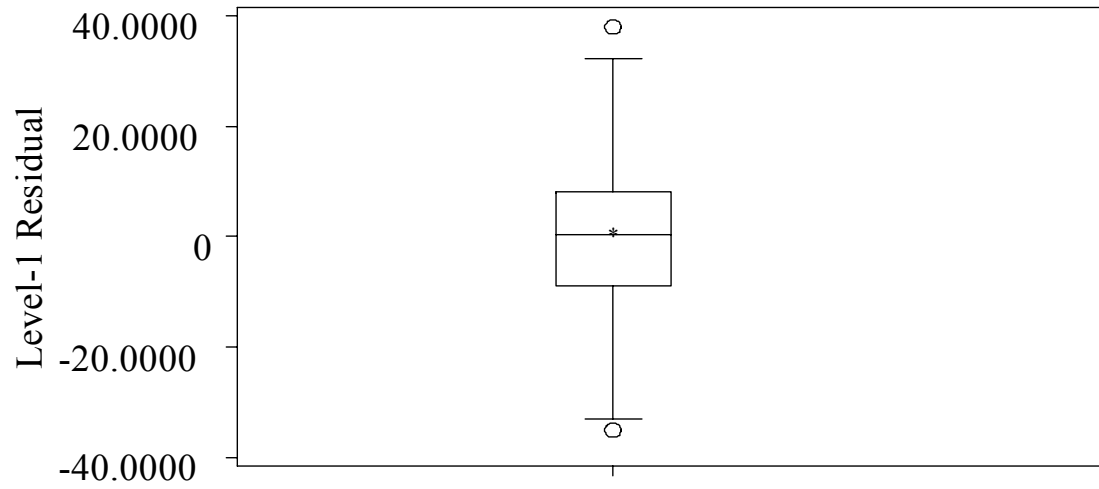


---

(sk = -0.18, ku = 2.87)

Figure H1. Box-and-whisker plot for level-1 residuals (instructional strategies).

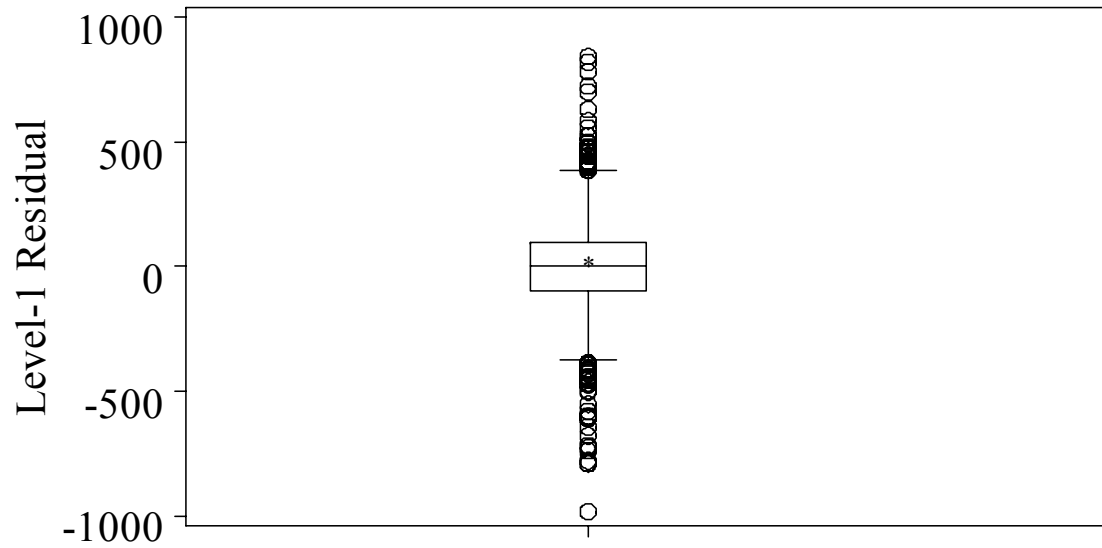
APPENDIX H (continued)



(sk = -0.01, ku = 0.19)

---

Figure H2. Box-and-whisker plot for level-2 residuals (instructional strategies).

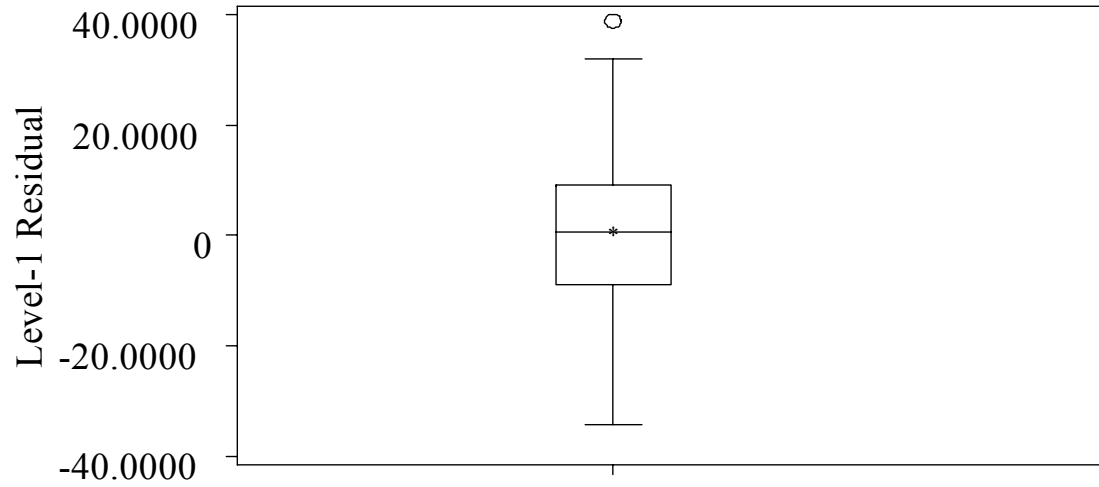


(sk = -0.17, ku = 2.85)

---

Figure H3. Box-and-whisker plot for level-1 residuals (classroom management).

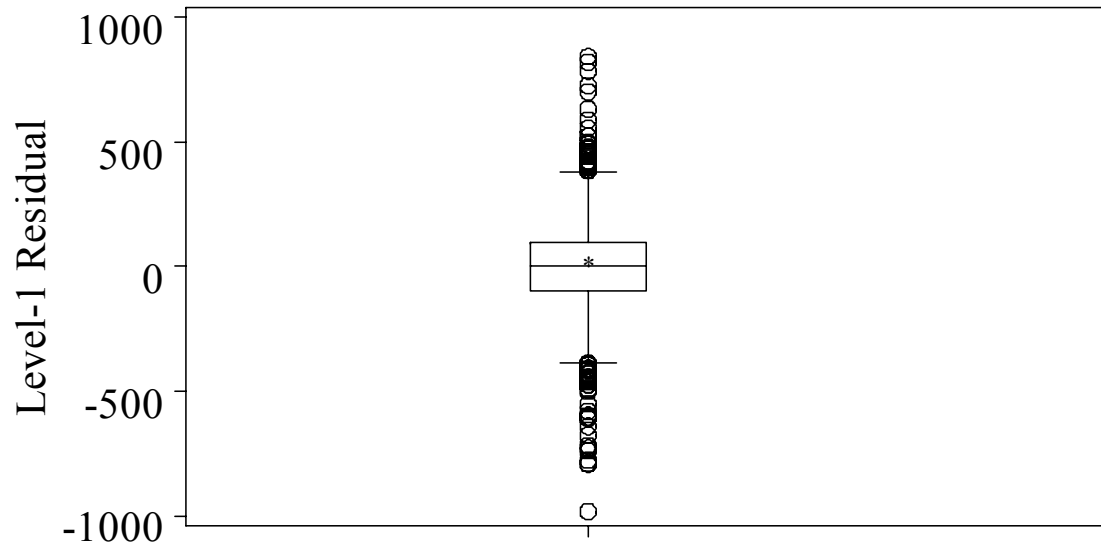
APPENDIX H (continued)



(sk = -0.00, ku = 0.17)

---

Figure H4. Box-and-whisker plot for level-2 residuals (classroom management).

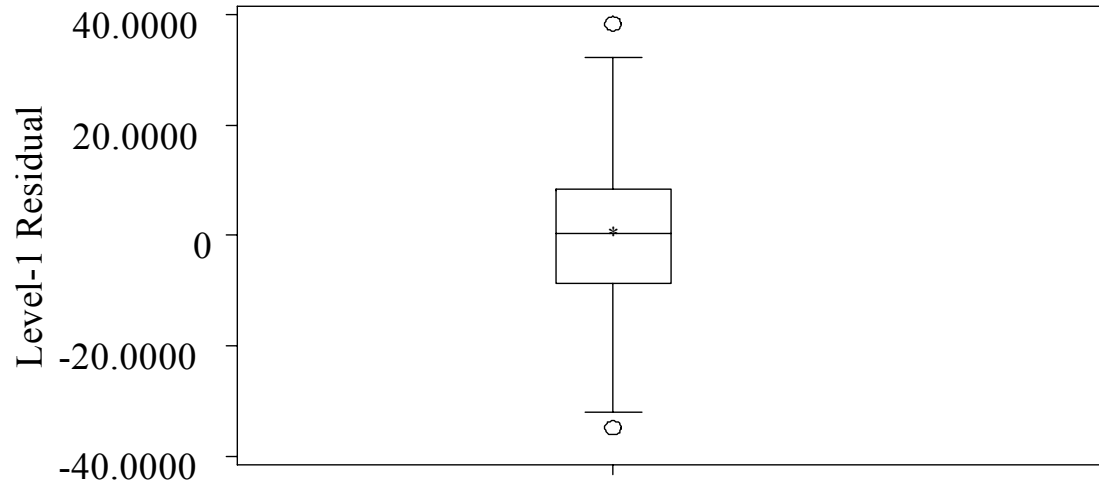


---

(sk = -0.18, ku = 2.85)

Figure H5. Box-and-whisker plot for level-1 residuals (student engagement).

APPENDIX H (continued)



(sk = 0.02, ku = 0.15)

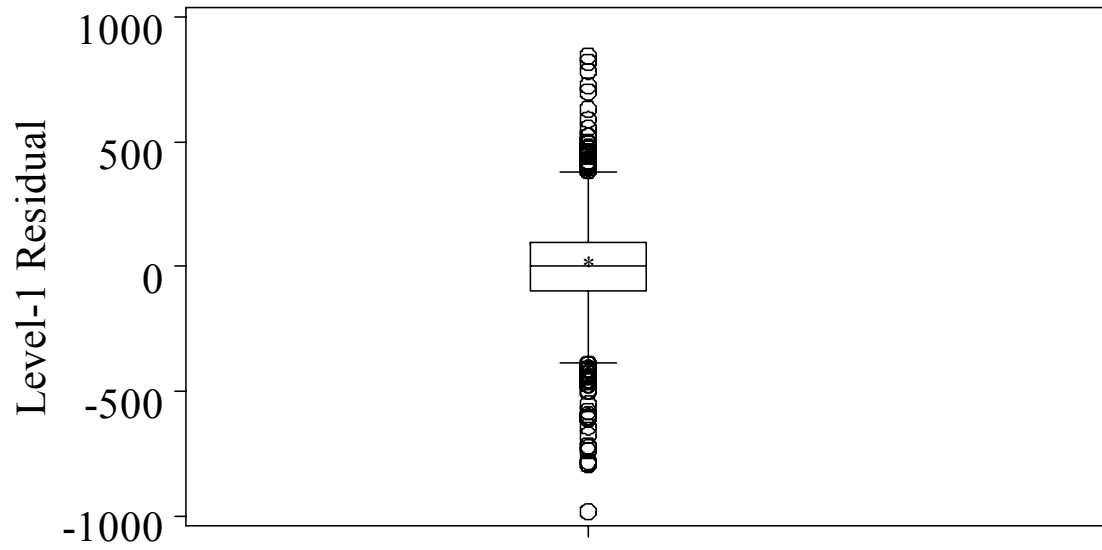
---

Figure H6. Box-and-whisker plot for level-2 residuals (student engagement).



APPENDIX H (continued)

---

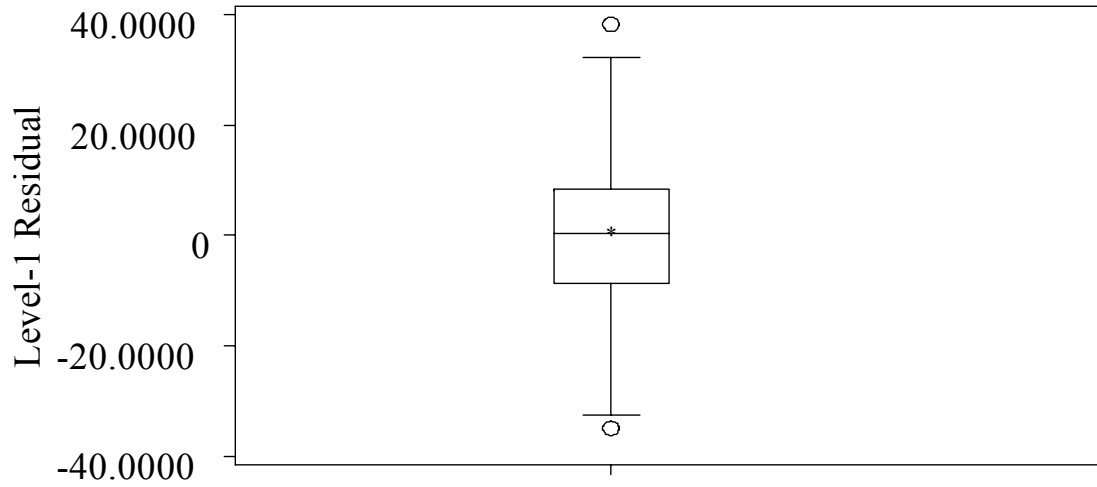


---

(sk = -0.18, ku = 2.86)

Figure H7. Box-and-whisker plot for level-1 residuals (total efficacy).

APPENDIX H (continued)



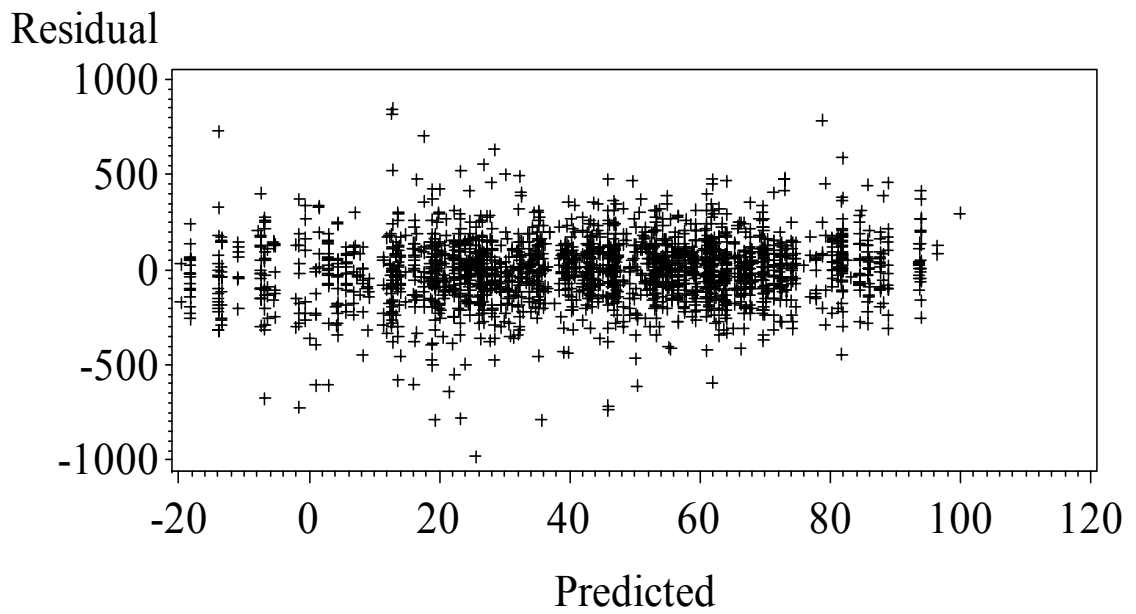
(sk = 0.01, ku = 0.17)

---

Figure H8. Box-and-whisker plot for level-2 residuals (total efficacy).

APPENDIX H (continued)

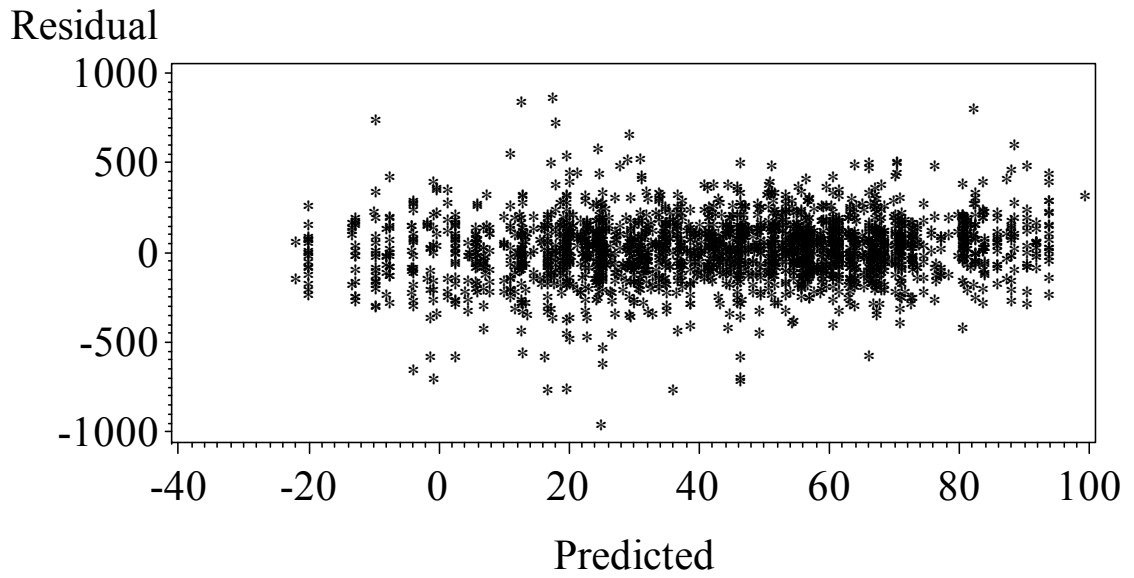
---



---

Figure H9. Level-2 Instructional Strategies residuals\*predicted reading achievement.

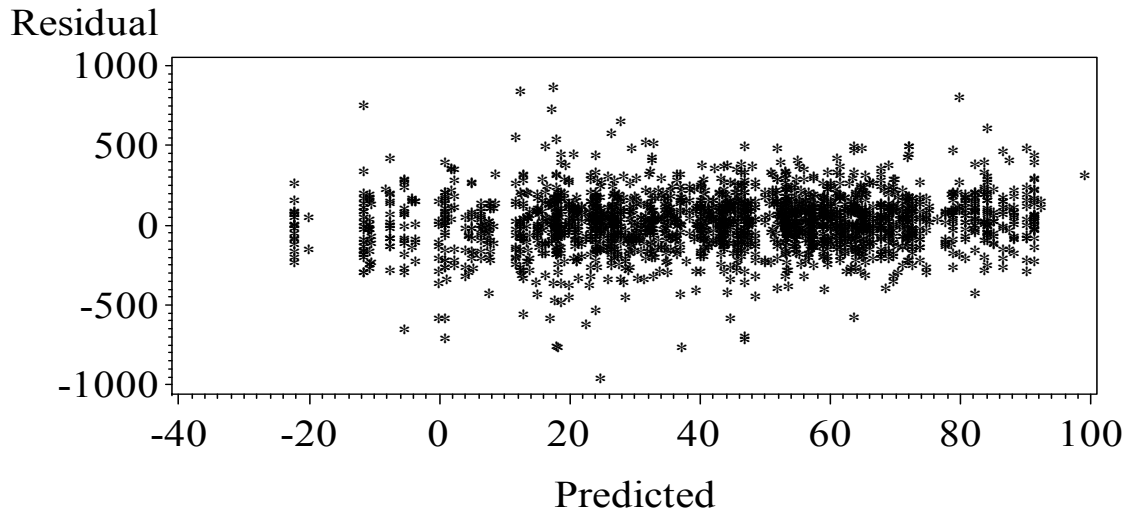
APPENDIX H (continued)



---

Figure H10. Level-2 Classroom Management residuals\*predicted reading achievement.

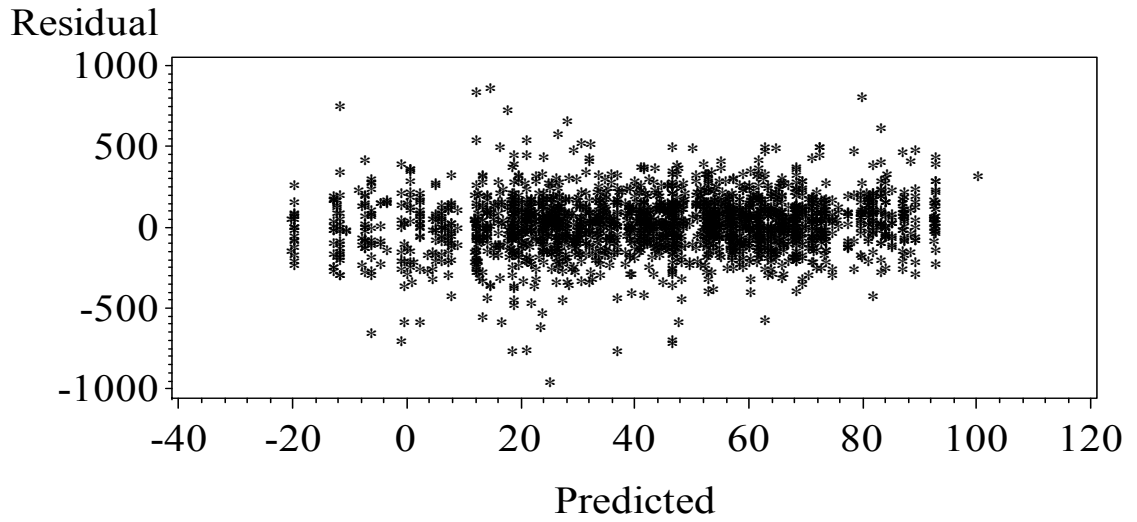
APPENDIX H (continued)



---

Figure H11. Level-2 Student Engagement residuals\*predicted reading achievement.

APPENDIX H (continued)



---

Figure H12. Level-2 Total Efficacy residuals\*predicted reading achievement.

APPENDIX I

Summary Tables for HLMs

Table I1

*HLMs for Instructional Strategies with Level-1 Variables*

	Model 2A/IS	Model 2B/IS	Model 2C/IS	Model 2D/IS
<i>Fixed Effects</i>				
Intercept	16.21 (39.96)	28.48 (37.84)	203.81 (105.53)	27.70 (38.71)
TotIS	0.49 (0.68)	0.52 (0.64)	0.42 (0.67)	0.44 (0.66)
Race		-31.89* (8.19)		
Grade			-19.48 (10.17)	
SES				-22.03* (8.21)
<i>Random Effects</i>				
Intercept, $\tau_{00}$	1210.95* (421.38)	895.06* (381.13)	1127.3* (411.41)	997.32* (397.76)
Residual, $\sigma^2$	30162* (968.87)	30148* (969.60)	30170* (969.15)	30198* (971.34)
<i>Model Fit</i>				
AIC	27163.3	27142.7	27153.2	27150.3
BIC	27168.7	27148.1	27158.6	27155.7

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses.

ICC of TotIS model = .039.

\*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I2

*HLMs for Instructional Strategies with Two-Way Main Effects and Interactions*

	Model 3A/IS	Model 3B/IS	Model 3C/IS	Model 3D/IS	Model 3E/IS	Model 3F/IS
<i>Fixed Effects- Main</i>						
Intercept	217.59* (101.07)	147.37 (119.12)	204.79* (99.22)	159.18 (121.65)	32.51 (37.67)	35.47 (37.72)
TotIS	0.38 (0.65)	0.37 (0.65)	0.46 (0.64)	0.47 (0.63)	0.49 (0.64)	0.48 (0.64)
Race			-31.66* (8.16)	68.28 (155.21)	-27.36* (9.02)	-36.10* (11.76)
Grade	-19.71* (9.73)	-12.20 (11.84)	-18.34 (9.57)	-13.54 (12.11)		
SES	-22.54* (8.18)	149.65 (155.47)			-11.02 (9.01)	-21.88 (13.01)
<i>Fixed Effects- Interactions</i>						
SES*Grade		-18.21 (16.42)				
Race*Grade				-10.58 (16.40)		
Race*SES						20.56 (17.76)
<i>Random Effects</i>						
Intercept, $\tau_{00}$	899.79* (386.56)	891.05* (385.83)	820.97* (372.03)	809.70* (370.98)	852.74* (376.15)	848.18* (374.81)
Residual, $\sigma^2$	30212* (972.01)	30214* (972.31)	30156* (969.85)	30172* (970.55)	30167* (970.62)	30165* (970.68)
<i>Model Fit</i>						
AIC	27139.9	27131.2	27132.7	27124.9	27135.0	27126.1
BIC	27145.9	27136.6	27138.1	27130.3	27140.4	27131.5

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .



APPENDIX I (continued)

Table I3

<i>HLMs for Instructional Strategies with Three-Way Main Effects and Interactions</i>		
	Model 3G/IS	Model 3H/IS
<i>Fixed Effects-</i>		
<i>Main</i>		
Intercept	211.86* (98.32)	216.84* (98.49)
TotIS	0.43 (0.63)	0.42 (0.63)
Race	-26.88* (9.01)	-34.60* (11.68)
Grade	-18.64* (9.47)	-18.89* (9.48)
SES	-11.63 (9.00)	-21.31 (12.96)
<i>Fixed Effects-</i>		
<i>Interaction</i>		
Race*Grade*SES		1.94 (1.86)
<i>Random Effects</i>		
Intercept ( $\tau_{00}$ )	770.36* (366.39)	772.79* (366.21)
Residual ( $\sigma^2$ )	30179* (971.08)	30176* (971.13)
<i>Model Fit</i>		
AIC	27124.9	27120.7
BIC	27130.3	27126.1

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I4

*HLMs for Classroom Management with Level-1 Variables*

	Model 2A/CM	Model 2B/CM	Model 2C/CM	Model 2D/CM
<i>Fixed Effects</i>				
Intercept	46.88 (40.82)	59.73 (38.67)	227.60* (101.22)	56.18 (39.42)
TotCM	-0.04 (0.67)	-0.02 (0.64)	0.08 (0.67)	-0.05 (0.65)
Race		-31.73* (8.19)		
Grade			-19.90* (10.22)	
SES				-22.12* (8.21)
<i>Random Effects</i>				
Intercept, $\tau_{00}$	1225.77* (422.97)	912.54* (383.12)	1139.73* (412.77)	1009.34* (399.08)
Residual, $\sigma^2$	30162* (968.76)	30147* (969.44)	30169* (969.03)	30197* (971.24)
<i>Model Fit</i>				
AIC	27163.8	27143.4	27153.6	27150.8
BIC	27169.2	27148.8	27159.0	27156.2

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses.

ICC of the TotCM model = .039. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I5

*HLMs for Classroom Management with Two-Way Main Effects and Interactions*

	Model 3A/CM	Model 3B/CM	Model 3C/CM	Model 3D/CM	Model 3E/CM	Model 3F/CM
<i>Fixed Effects- Main</i>						
Intercept	238.64* (96.91)	165.83 (116.45)	230.16* (95.30)	184.80 (118.76)	62.60 (38.40)	66.67 (38.51)
TotCM	0.08 (0.64)	0.09 (0.64)	0.09 (0.63)	0.12 (0.63)	-0.03 (0.63)	0.06 (0.63)
Race			-31.53* (8.17)	67.33 (155.38)	-27.14* (9.02)	-36.00* (11.76)
Grade	-20.08* (9.78)	-12.49 (11.88)	-18.79* (9.63)	-14.06 (12.14)		
SES	-22.62* (8.18)	151.77 (155.53)			-11.22 (9.01)	-22.22 (13.01)
<i>Fixed Effects- Interactions</i>						
SES*Grade		-18.44 (16.42)				
Race*Grade				-10.47 (16.42)		
Race*SES						20.82 (17.78)
<i>Random Effects</i>						
Intercept, $\tau_{00}$	909.44* (387.73)	900.05* (386.89)	835.83* (373.78)	825.00* (372.75)	867.89* (377.92)	862.76* (376.43)
Residual, $\sigma^2$	30211* (971.92)	30213* (972.23)	30154* (969.70)	30170* (970.39)	30167* (970.50)	30164* (970.54)
<i>Model Fit</i>						
AIC	27140.2	27131.5	27133.3	27125.4	27135.6	27126.7
BIC	27145.6	27136.9	27138.7	27130.8	27141.0	27132.1

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I6

*HLMs for Classroom Management with Three-Way Main Effects and Interactions*

	Model 3G/CM	Model 3H/CM
<i>Fixed Effects-</i>		
<i>Main</i>		
Intercept	235.62* (94.34)	241.44* (94.55)
TotCM	0.09 (0.62)	0.06 (0.62)
Race	-26.68* (9.01)	-34.45* (11.67)
Grade	-19.06* (9.53)	-19.27* (9.54)
SES	-11.81 (9.00)	-21.54 (12.96)
<i>Fixed Effects-</i>		
<i>Interaction</i>		
Race*Grade*SES		1.95 (1.87)
<i>Random Effects</i>		
Intercept ( $\tau_{00}$ )	782.85* (367.95)	785.34* (367.72)
Residual ( $\sigma^2$ )	30177* (970.98)	30174* (971.01)
<i>Model Fit</i>		
AIC	27125.3	27121.2
BIC	27130.7	27126.6

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I7

*HLMs for Student Engagement with Level-1 Variables*

	Model 2A/SE	Model 2B/SE	Model 2C/SE	Model 2D/SE
<i>Fixed Effects</i>				
Intercept	6.04 (29.26)	29.19 (28.38)	183.87 (107.42)	22.17 (28.94)
TotSE	0.74 (0.55)	0.56 (0.53)	0.54 (0.56)	0.59 (0.54)
Race		-31.03* (8.22)		
Grade			-17.77 (10.33)	
SES				-21.28* (8.25)
<i>Random Effects</i>				
Intercept, $\tau_{00}$	1164.80* (415.99)	882.61* (379.24)	1106.62* (409.19)	975.75* (394.56)
Residual, $\sigma^2$	30170* (969.28)	30149* (969.60)	30174* (969.36)	30200* (971.39)
<i>Model Fit</i>				
AIC	27163.8	27142.7	27153.0	27149.9
BIC	27169.2	27148.1	27158.4	27155.3

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses.

ICC of TotSE model = .037. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I8  
*HLMs for Student Engagement with Two-Way Main Effects and Interactions*

	Model 3A/SE	Model 3B/SE	Model 3C/SE	Model 3D/SE	Model 3E/SE	Model 3F/SE
<i>Fixed Effects- Main</i>						
Intercept	208.17* (103.55)	138.01 (121.23)	202.39* (101.64)	157.27 (123.53)	35.06 (28.48)	36.81 (28.55)
TotSE	0.39 (0.54)	0.38 (0.54)	0.37 (0.35)	0.37 (0.53)	0.51 (0.52)	0.50 (0.52)
Race			-31.05* (8.20)	68.11 (155.24)	-26.73* (9.03)	-35.46* (11.77)
Grade	-18.53 (9.92)	-11.06 (11.98)	-17.30 (9.76)	-12.53 (12.27)		
SES	-22.02* (8.23)	150.02 (155.44)			-10.60 (9.03)	-21.44 (13.02)
<i>Fixed Effects- Interactions</i>						
SES*Grade		-18.19 (16.41)				
Race*Grade				-10.50 (16.40)		
Race*SES						20.52 (17.76)
<i>Random Effects</i>						
Intercept, $\tau_{00}$	897.75* (385.92)	889.86* (385.18)	824.57* (372.28)	813.93* (371.21)	844.99* (374.75)	840.88* (373.46)
Residual, $\sigma^2$	30211* (971.90)	30213* (972.17)	30154* (969.73)	30170* (970.41)	30167* (970.56)	30165* (970.60)
<i>Model Fit</i>						
AIC	27140.1	27131.4	27133.2	27125.3	27135.1	27126.1
BIC	27145.5	27136.8	27138.6	27130.7	27140.5	27131.5

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I9

*HLMs Student Engagement with Three-Way Main Effects and Interactions*

	Model 3G/SE	Model 3H/SE
<i>Fixed Effects-</i>		
<i>Main</i>		
Intercept	212.36* (100.94)	217.17* (101.09)
TotSE	0.31 (0.53)	0.31 (0.53)
Race	-26.43* (9.02)	-34.20* (11.69)
Grade	-17.78 (9.67)	-18.04 (9.68)
SES	-11.39 (9.03)	-21.11 (12.98)
<i>Fixed Effects-</i>		
<i>Interaction</i>		
Race*Grade*SES		1.95 (1.86)
<i>Random Effects</i>		
Intercept ( $\tau_{00}$ )	776.97* (366.93)	779.22* (366.72)
Residual ( $\sigma^2$ )	30176* (970.91)	30173* (971.96)
<i>Model Fit</i>		
AIC	27125.3	27121.2
BIC	27130.7	27126.6

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I10

*HLMs for Total Efficacy with Level-1 Variables*

	Model 2A/TE	Model 2B/TE	Model 2C/TE	Model 2D/TE
<i>Fixed Effects</i>				
Intercept	10.94 (41.72)	29.68 (39.72)	198.61 (107.25)	25.59 (40.57)
TotTE	0.20 (0.24)	0.17 (0.23)	0.16 (0.24)	0.16 (0.23)
Race		-31.59* (8.20)		
Grade			-19.29 (10.17)	
SES				-21.85* (8.22)
<i>Random Effects</i>				
Intercept, $\tau_{00}$	1202.69* (420.75)	896.71* (381.48)	1122.66* (411.15)	994.90* (397.59)
Residual, $\sigma^2$	30165* (969.02)	30148* (969.62)	30172* (969.25)	30199* (971.40)
<i>Model Fit</i>				
AIC	27163.8	27144.9	27155.1	27152.3
BIC	27169.2	27150.3	27160.5	27157.7

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses.

ICC for TotTE model = .038. \*Statistically significant – variance estimate and intercept,  $p < .05$ .



APPENDIX I (continued)

Table I11  
*HLMs for Total Efficacy with Two-Way Main Effects and Interactions*

	Model 3A/TE	Model 3B/TE	Model 3C/TE	Model 3D/TE	Model 3E/TE	Model 3F/TE
<i>Fixed Effects- Main</i>						
Intercept	216.40* (102.93)	145.09 (120.96)	207.01* (101.07)	161.07 (123.36)	34.42 (39.60)	37.88 (39.67)
TotTE	0.13 (0.23)	0.13 (0.23)	0.13 (0.23)	0.14 (0.23)	0.16 (0.23)	0.15 (0.23)
Race			-31.40* (8.17)	68.80 (155.26)	-27.11* (9.02)	-35.81* (11.76)
Grade	-19.59* (9.75)	-12.02 (11.85)	-18.25 (9.60)	-13.43 (12.13)		
SES	-22.39* (8.20)	150.98 (155.45)			-10.95 (9.02)	-21.77 (13.02)
<i>Fixed Effects- Interactions</i>						
SES*Grade		-18.34 (16.41)				
Race*Grade				-10.61 (16.41)		
Race*SES						20.46 (17.77)
<i>Random Effects</i>						
Intercept, $\tau_{00}$	900.30* (386.68)	891.37* (385.88)	825.11* (372.63)	813.89* (371.54)	855.17* (376.56)	851.15* (375.28)
Residual, $\sigma^2$	30213* (972.02)	30214* (972.31)	30155* (969.82)	30171* (970.52)	30168* (970.63)	30165* (970.68)
<i>Model Fit</i>						
AIC	27142.0	27133.3	27135.0	27127.1	27137.2	27128.3
BIC	27147.4	27138.7	27140.4	27132.5	27142.6	27133.7

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

APPENDIX I (continued)

Table I12

*HLMs Total Efficacy with Three-Way Main Effects and Interactions*

	Model 3G/TE	Model 3H/TE
<i>Fixed Effects- Main</i>		
Intercept	215.07* (100.22)	220.64* (100.42)
TotTE	0.12 (0.23)	0.12 (0.23)
Race	-26.65* (9.01)	-34.45* (11.69)
Grade	-18.57* (9.49)	-18.83* (9.50)
SES	-11.59 (9.01)	-21.24 (12.97)
<i>Fixed Effects- Interaction</i>		
Race*Grade*SES		1.93 (1.86)
<i>Random Effects</i>		
Intercept ( $\tau_{00}$ )	775.09* (367.03)	777.83* (366.88)
Residual ( $\sigma^2$ )	30178* (971.05)	30175* (971.09)
<i>Model Fit</i>		
AIC	27127.1	27122.9
BIC	27132.5	27128.3

*Note:* Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses. \*Statistically significant – variance estimate and intercept,  $p < .05$ .

## APPENDIX J

### An Explanation of Developmental Scale Scores

DSS scores are created so that students' test scores can be comparable from year to year. Scores are originally reported as standardized scale scores (SSS) on the FCAT 100–500 scale. To make SSSs comparable to scores from past years, the data output by Item Response Theory (IRT) processing is equated (Lord & Novick, 1968). This equating process involves (a) repeating in the 2007 test “anchor items” that were used in previous FCAT administrations, and (b) applying the Stocking/Lord (1983) procedure using parameters from those anchor items to adjust for the difference between students in 2007 and previous years. The anchor items and the Stocking/Lord procedure are used to equate 2007 test scores to the test scores originally reported. This procedure, with different anchor items, has been conducted every year since 1998.

With the completion of the 2007 scaling, the anchor items have two sets of item parameters: (a) new parameters on the mean equal to 0 and standard deviation equal to 1 scale produced this year, and (b) old parameters that were transformed during their previous use. The old parameters are based on either the original 1998 scale or the 2001 scale. The Stocking/Lord (1983) procedure uses the old item parameters to locate the achievement scale and then searches for a transformation multiplier and additive constant that can combine to make the new parameters replicate the original achievement scale as closely as possible. This is done by attempting to match test characteristic curves (TCC), which are summations of item characteristic curves (ICC), produced by the old

parameters with TCCs produced by transformations of new parameters. Since the items are the same, the same scale should result.

During this equating process, item-level reviews are conducted. Specifically, item parameter estimates are reviewed for their stability before they are included in the equating process. A tool used to evaluate anchor parameter differences is a computational procedure that produces a metric indicating the difference between the shapes of the ICCs produced by the current item parameters versus base-year item parameters, parameters that are equated to the base scale in the most recent administration of the items. This metric takes all item parameters into account. The procedure checks for outlier items by computing differences in response probabilities based on base-year and current year parameter values. The items with the largest differences are identified for further review and possible elimination from equating. In order to calculate the differences, anchor parameters are placed on the current year's IRT scale. Then absolute values of the differences are calculated at 31 quadrature points, the same that are used in the Stocking/Lording procedure, and the mean of the 31 absolute differences is computed for each item. Items are flagged if their mean absolute difference is greater than expected, given the mean absolute differences of all items. If a particular item parameter is too low, too high, or at variance with prior parameter estimates, then Florida Department of Education (FDOE) personnel make a decision as to whether the item should remain in the anchor set.

Another method used to compare old and new item parameter differences is to plot two item characteristic curves for each anchor item: one plot is created by using the

previous year's parameters, and the second is created using the current year's parameters. This allows for visual comparison of the two ICCs. This technique adds another useful decision-making tool to those that are already in place.

Anchor items should have as many of the relevant characteristics, to the extent possible, as the core items. Several statistical comparisons are done to examine this issue. First, a comparison of the mean proportion correct is calculated. For instance, the mean for core items answered correctly is compared to the mean for anchor items answered correctly. Another statistic used to compare anchor and core item behavior is mean points scored for core items versus anchor items. Total points from anchor items should be at least 20 % of the total points scored on the core test. These comparisons are made by content category as well as by item type (multiple-choice, gridded-response, short response, and extended response).

IRT scaling algorithms attempt to find item parameters, or numerical characteristics, that create a match between observed patterns of item responses and theoretical response patterns defined by the selected IRT models. The Q1 statistic (Yen, 1981) is used as an index for finding how well theoretical item curves match observed item responses. Q1 is computed by first conducting an IRT item-parameter estimation, then by estimating students' achievement using the estimated item parameters, and lastly, by using students' achievement scores in combination with estimated item parameters to compute expected performance on each item. Differences between expected item performance and observed item performance are then compared at selected intervals across the range of student achievement. Q1 is computed as a ratio involving expected

and observed item performance and is therefore interpretable as a chi-square statistic which is a statistical test that determines whether the data, observed item performance, fits the hypothesis, the expected item performance.

Q1 for each item type has varying degrees of freedom because the different types of items have different numbers of IRT parameters. Therefore, Q1 is not directly comparable across item types. An adjustment, or linear transformation, translation to a z-score, ZQ1, is made for different numbers of item parameters and sample size to create a more comparable statistic. The FCAT has set a criteria for a minimum ZQ1 value standard for an item to have acceptable fit (FDOE, 1998).

By fitting all items simultaneously to the same achievement scale, IRT operates under the assumption that there is a strong, single construct that underlies the performance of all items. Under this assumption, performance on the items should be related to achievement, and additionally, any relationship of performance between pairs of items should be accounted for by variance in students' levels of achievement. This is the "local dependence" assumption of unidimensional IRT and suggests a relatively straightforward test for unidimensionality, called the Q3 statistic (Yen, 1984).

To compute the Q3 statistic, expected student performance on each item is calculated using item parameters and estimated achievement scores. Then for each student and each item, the difference between expected and observed item performance is calculated. The difference can be thought of as: what is left in performance after accounting for underlying achievement? If performance on an item is driven by a single achievement construct, then not only will the residual be small, but the correlation

between residuals of the pair of items will also be small. These correlations are analogous to partial correlations, which can be interpreted as the relationship between two variables, or items, after the effects of a third variable, underlying achievement, are held constant. The correlation among IRT residuals is the Q3 statistic. These data, coupled with the Q1 data above, indicate that the unidimensional IRT model provides a very reasonable solution for capturing the essence of student achievement defined by the carefully selected set of items for each grade and subject.

## ABOUT THE AUTHOR

Anete Vásquez received her Bachelor of Arts degree in English and her Master's of Education degree from the University of Florida. She taught English for eleven years and spent her last five years teaching in the International Baccalaureate Program. She was a regional finalist for Florida Teacher of the Year and won a Milken Family Foundation National Teacher Award.

Ms. Vásquez is currently the Coordinator of Field Experiences for the Department of Secondary Education at the University of South Florida and teaches methods courses and classroom management. She has won both the Provost's Award for Outstanding Teaching by a Graduate Assistant and a Faculty Outstanding Undergraduate Educator Award. She has presented at the American Educational Research Association and the National Council of Teachers of English and is currently working on a book about English language learners in the English/Language Arts classroom.