Abstract

This quantitative study focuses on the relationship between classroom quality and children’s academic achievement. Specifically, it examines how classroom quality in three broad domains—emotional climate, classroom management and instructional support—impact kindergarten achievement growth in mathematics and reading. The researcher collected data from 28 kindergarten classes in a small mid-western school district. These data include measures of classroom quality using the Classroom Assessment Scoring System (CLASS) and measures of student achievement using the Northwest Evaluation Association Measures of Academic Progress (NWEA MAP). Analysis of the relationship between classroom quality and student achievement is mixed, providing inconclusive results. While mathematics growth scores during the first growth period show a significant positive relationship to classroom quality, reading growth scores are not significantly related during this growth period. Neither winter to spring growth scores nor overall growth scores show a significant relationship to classroom quality in this study. The researcher recommends further research to determine why outcomes in this study did not support findings by other researchers.
ACKNOWLEDGEMENTS

The researcher wishes to acknowledge and thank Dr. Karen Ford, Dr. Patricia Clark, Dr. Cheryll Adams, Dr. Nancy Melser, Dr. Holmes Finch and Dr. Dale Umbach for their guidance, expertise and encouragement, without which this dissertation would not have been possible.

Appreciation is also extended to the administrators and teachers in the study school district for allowing me to observe and complete my research in their schools and classrooms.

The researcher is grateful for the support and encouragement of her colleagues in the Department of Education at Huntington University; their support and encouragement have been invaluable.

A special word of appreciation is extended to Scott, Lindsey, Ashley and Ryan Burson for their love and sacrifice during the time of this study and research. In addition, she owes a debt of gratitude to Mr. and Mrs. B. E. Heise, Jr. for teaching her the value of hard work and believing she is capable of accomplishing whatever she wants in life.

SJB
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ............................................................................................................. v  

LIST OF TABLES AND FIGURES ........................................................................................... vii  

I. INTRODUCTION ...................................................................................................................... 9  
   Problem Studied .......................................................................................................................... 11  
   Research Questions ..................................................................................................................... 12  
   Data Collection .......................................................................................................................... 12  
   Definition of Terms ................................................................................................................... 15  
   Significance of Problem .......................................................................................................... 16  
   Basic Assumptions .................................................................................................................... 17  
   Basic Limitations ....................................................................................................................... 18  

III. LITERATURE REVIEW ........................................................................................................... 20  
   Historical Background ............................................................................................................. 20  
   Defining Quality ....................................................................................................................... 21  
   Effects of Early Childhood Programs ....................................................................................... 23  
   Quality Variables in Elementary School .................................................................................. 40  
   Quality in Kindergarten ........................................................................................................... 46  
   Classroom Assessment Scoring System (CLASS) .................................................................. 55  

III. METHODOLOGY .................................................................................................................... 57  
   Purpose ..................................................................................................................................... 57
Setting 57
Participants 58
Measures 61
Procedure 75

IV. RESULTS ........................................................................................................... 82

Overview 82
Research Questions 83
Descriptive Statistics for CLASS Domains and Dimensions 83
Descriptive Statistics for NWEA MAP 86

V. CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS ................. 98

Analysis of Data 98
Research Questions and Hypotheses 101
Discussion 102
Recommendations for Future Research and Action 114
Summary and Conclusions 117

REFERENCES ........................................................................................................ 118

APPENDICES ..................................................................................................... 139

Appendix A – Teacher Survey 140
Appendix B – Summary of CLASS Domains and Dimensions 142
# LIST OF TABLES AND FIGURES

<table>
<thead>
<tr>
<th>Figure/Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3.1</td>
<td>Overview of Class Domains and Dimensions</td>
<td>65</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Descriptive Statistics of CLASS</td>
<td>84</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Descriptive Statistics for NWEA MAP Standard Scores in Mathematics and Reading</td>
<td>88</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Descriptive Statistics for NWEA MAP Gain Scores in Mathematics and Reading</td>
<td>90</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Significance of Classroom Quality – Slope Coefficients</td>
<td>96</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Estimated Marginal Means for Significant Fixed Effects</td>
<td>97</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Descriptive Statistics ((M, SD and Range)) on CLASS across Six Studies – Comparison to Current Study</td>
<td>100</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Lifelong Guidelines and LIFESKILLS</td>
<td>110</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

As professionals and government officials in United States society become increasingly focused on academic achievement and accountability in education, mounting evidence of the lasting value of quality early childhood programs emerges (Barnett, 1995, 1998; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Gorey, 2001; La Paro, Pianta, & Stuhlman, 2004; NAEYC, 1997; Reeves, 2004; Schweinhart, 2005; Siraj-Blatchford, Sylva, Muttock, Gilden, & Bell, 2002; Weikart, 2000). Because of this evidence, increasing attention is directed toward early childhood programs and the roles they may play in setting the stage for student success in later grades. The age-old debate regarding what constitutes quality and best practice resurfaces with particular attention given to classroom quality and its long-term effects on students. Government officials and educators continue to ask what can be done to improve student outcomes through implementation of new curricula, modification of the environment, and additional assessment. The emphasis is on accountability for learner achievement.

The No Child Left Behind Act of 2001 [NCLB], signed into law by President George W. Bush, is a major influence initiating recent changes in the American education system. This legislation, which “reauthorized the Elementary and Secondary Education Act of 1964 (ESEA) – the principal federal law affecting education from kindergarten through high school” addresses four main foci of education: “accountability for results, an emphasis on doing what works based on scientific research, expanded parental options, and expanded local control and flexibility” in implementation of programs and
systems of accountability (U. S. Department of Education, 2004, p.1). The act “highlights the importance of quality in early education classrooms and emphasizes accountability and the use of [research-based] teaching methods that improve student outcomes” (La Paro et al., 2004, p. 410). However, in spite of this pressing mandate and extensive examination of quality in child care settings, relatively little research has been conducted on classroom quality in public elementary classrooms (Pianta, La Paro, Payne, Cox, & Bradley, 2002).

For more than four decades, research has been conducted addressing the effects of early childhood programs, with results showing positive long-term benefits for children enrolled in quality child care and preschool (Weikart, 2000). Because of the above issues and a current focus on accountability and student outcomes, additional research has been employed to assess specific variables related to student success, particularly the impact of classroom and teacher quality on achievement. Past studies conceptualize and evaluate early childhood program quality in a variety of ways: in terms of class size, curriculum selection, materials, space for play, student safety, teacher credentials and length of school day, as well as whether instructional practices are deemed appropriate for children at varying stages of development (Pianta et al., 2002a; Zahorik, 1999). In addition, these studies define learner outcomes in terms of social, emotional, intellectual and physical development of students (Weikart, 2000).

In this study, specific elements of classroom quality were identified and compared to academic achievement in mathematics and reading for kindergarten students in a small mid-western school district. The Classroom Assessment Scoring System (CLASS) was used to identify and evaluate specific classroom-level variables to establish global quality
ratings for each classroom in three domains: emotional support, classroom organization, and instructional support (Pianta, La Paro, & Hamre, 2008). To measure academic achievement, the Northwest Evaluation Association Measures of Academic Progress (NWEA MAP) was used. This assessment “include[s] Early Literacy and Early Numeracy Screening (diagnostic) tests [and] Skills Checklist (diagnostic) tests” in addition to Survey with Goals (adaptive) tests in Reading and Mathematics (Northwest Evaluation Association: Overview, 2008).

Problem Studied

The school district in which this study took place had approximately 450 kindergarten students in eight elementary schools during the study period. This district, like most, is dealing with pressure from the state to meet NCLB mandates “to achieve unprecedented educational progress…ensuring that all students—and all subgroups—meet the state’s proficiency goals” (Sunderman & Orfield, 2007, p. 138). These concerns must be addressed as teachers strive to prepare students for standardized testing and readiness for future grades while balancing academics with developmentally appropriate practice. To address these issues, classroom observations and analyses were conducted with the following goal in mind; to assess the relationship between classroom quality and kindergarten achievement. The researcher examined ten dimensions of classroom quality using a standardized assessment instrument (CLASS). Student academic achievement in mathematics and reading, as measured by NWEA MAP, was compared with global classroom quality ratings and specific quality indicators to determine if classroom quality and student achievement in mathematics and reading were related.
Research Questions

Question 1: Is kindergarten achievement growth in mathematics related to global classroom quality as measured by the CLASS?

Question 2: Is kindergarten achievement growth in reading related to global classroom quality as measured by the CLASS?

Question 3: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in mathematics?

Question 4: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in reading?

Data Collection

The school district assessed all students entering kindergarten during the first weeks of the academic year to establish baseline achievement levels in mathematics and reading using the Northwest Evaluation Association Measures of Academic Progress (NWEA MAP), Primary Version. “[These tests] are electronically administered and scored achievement tests designed to measure growth in student learning for individual students, classrooms, schools, and districts” (Northwest Evaluation Association: FAQ, 2008).

After attaining baseline achievement scores for all kindergarten children in the district, the researcher assessed global classroom quality for each kindergarten class using the Classroom Assessment Scoring System (CLASS). The CLASS “provides a framework for observing key dimensions of classroom processes…that contribute to quality of the classroom setting from preschool through third grade” (La Paro et al., 2004, p. 409). The researcher collected four 20- to 30-minute cycles of data in each class, with
the exception of one class in which the teacher went on extended leave prior to completion of data collection. The researcher calculated an average quality score for each of the 10 dimensions observed during data collection. These dimensions fall into three broader domains:

1. Emotional Support – which includes elements of positive climate, negative climate, teacher sensitivity, and regard for student perspectives.
2. Classroom Organization – which includes elements of behavior management, productivity, and instructional learning formats.
3. Instructional Support – which includes elements of concept development, quality of feedback, and language modeling. (Pianta et al., 2008)

The CLASS Manual requires the observer to score for each quality dimension during each observation cycle. These scores are based on the extent to which certain behavioral markers characterize the classroom during that cycle (Pianta et al., 2008). The researcher conducted four observation cycles in each district kindergarten class, with the exception of one class in which only two cycles were recorded. She then averaged all cycles for each classroom to derive a mean score on each dimension for each class.

To develop accuracy in scoring and understanding necessary for interpretation of observations, the researcher participated in CLASS reliability training provided by developers of the instrument at the University of Virginia, Charlottesville. After explicit training to establish understanding of the purposes and procedures associated with the instrument, multiple videotaped segments of actual classroom sessions were viewed to reinforce understanding of the dimensions, as well as observable indicators of each. After training and practice sessions, the researcher was tested. Several videotaped classroom
sessions were viewed and scored, then compared to the standard for accuracy. The researcher established an 80 percent reliability rating as a scorer.

The investigator used random cluster sampling by school to determine the order in which observations took place. She placed all school names in a container and drew a school name for each day of observation. If the selected school could not be visited on the date selected, the researcher drew another school name. Reasons to reject an initial selection included: all observations were complete at a specific school, the class was not in session during the observation period (at lunch, at related arts, on a field trip) or the teacher was absent from the classroom. Once in a school, the researcher conducted observations in as many kindergarten classes in the building as time permitted.

After classroom observation and assessment were completed, the researcher received post-test achievement scores for the NWEA MAP in reading and mathematics from the district’s assistant superintendent. She compared baseline reading and math (fall scores) to retest scores (winter and spring) for the NWEA MAP to determine academic growth over the course of the study. In addition, she coded for economic status (by full-pay, free, or reduced meal price designation), gender, school, and teacher variables, so that these demographics would be accounted for in the statistical analysis of results. Finally, the researcher compared scores on the CLASS to academic growth scores on NWEA MAP in mathematics and reading to determine if classroom quality is related to achievement in mathematics and reading.
Definition of Terms

1. Achievement growth – “the difference in [achievement test] scores for a single student from one point in time to another” (Cronin, Kingsbury, McCall, & Bowe, 2005, para. 3)

2. Achievement level – “the score [on an achievement test] that a student has at one point in time” (Cronin, Kingsbury, McCall, & Bowe, 2005, para. 3)

3. Alternating-day Kindergarten – Alternating-day kindergarten programs are defined as programs that meet all day on Monday and Wednesday or Tuesday and Thursday, and meet a half-day (morning or afternoon) on Friday.

4. Best practices – Best practices are those practices that draw upon research to promote optimal student outcomes in physical, social, emotional and cognitive development.

5. Child care – “care given for children by someone other than their parents and family members” (Segal, Bardige, Woika, & Leinfelder, 2006, p. 2).

6. Class – A class is defined as all children taught by the same teacher and who share the same school schedule.

7. Classroom Quality – is a measurement of structural and/or process characteristics of a classroom that affect outcomes for children (Pratt, as cited in Galley, 2000). Quality as measured by the Classroom Assessment Scoring System includes measures of emotional support, classroom management, and instructional support in the classroom setting (Pianta et al., 2008).

8. Curriculum – “At its simplest, curriculum is defined as what to teach and how to teach it” (Frede & Ackerman, 2007, p. 2).
9. Developmentally Appropriate Practice – Developmentally Appropriate Practice is defined as “teaching children in ways that meet children where they are, as individuals and as a group; and help each child reach challenging and achievable goals that contribute to his or her ongoing development and learning” (Copple & Bredekamp, 2006, p. 3; NAEYC, 1997).

10. Early Childhood Education – the schooling of children from birth to age eight (Bredekamp & Copple, 1997).

11. Elementary Education – the schooling of children typically in kindergarten through fifth grade.

12. Full-day Kindergarten – Full-day kindergarten is defined as a “program that meets for 4.5 to 6 hours per day, five days per week, and follows the same school calendar as early primary grades.” Weekly attendance in full-day programs averages 32 hours per week (Ackerman, Barnett, & Robin, 2005, p. 3).

13. Half-day Kindergarten – Half-day kindergarten is defined as a program that meets between two and three hours per day, five days per week. Weekly attendance in half-day programs averages 16 hours per week (Ackerman et al., 2005).

Significance of Problem

Along with expectations established by implementation of the No Child Left Behind Act, teachers, schools, and districts face increasing pressure to meet annual yearly progress (AYP) goals to demonstrate program quality. As these standards are imposed and administrators seek to initiate improvement, it is critical that implementation of change is based on research and empirical evidence rather than on individual preference, tradition, or program popularity (Slavin, 1989b). Teachers and administrators want
evidence that supports practices and programs that will have the greatest potential to improve student outcomes. The findings of this study identify indicators associated with student achievement over which teachers have control, thereby giving stakeholders additional information to guide decisions regarding how to improve classroom quality. The end goals are improved classroom quality and increased long-term achievement for all students in the district.

Although the immediate goals of this study relate to local improvement, the broader goal is to add to the professional knowledge base regarding the relationship of classroom and teacher quality to student achievement, particularly at the kindergarten level. In this age of accountability, it is especially valuable, and perhaps crucial, that districts conduct on-going rigorous self-assessment of all factors which contribute to student outcomes if “best practice” is to be identified and implemented in education settings.

Basic Assumptions

1. Standardization – All standardized tests were administered following guidelines for standardization.
2. Inclusion of scores – All student scores were included in data sets provided to the researcher.
3. Full participation – All district kindergarten classes and all eligible students participated in the study. Non-eligible students include students with incomplete data sets due to moving or absence and those who were identified as having special needs and were exempt from testing.
4. Coding of data – The district removed student names from data to maintain participant anonymity. Student scores and demographics were matched using identification numbers to which the researcher had no access. These were submitted to the researcher in spreadsheet format.

5. Willing participation – All kindergarten teachers in the district participated voluntarily and without coercion from the district or the researcher.

6. Sharing of findings – Aggregated research findings will be made available to teacher participants, participating schools, and the district administration upon completion and final approval of the study.

Basic Limitations

1. NWEA validity – There is some discussion in the field about the validity of any form of standardized assessment for children at such a young age.

2. Selection process – Differences in class make-up may affect study outcomes, as students were not placed randomly with teachers, but were placed due to socio-economic and at-risk factors. Students with at-risk factors received all-day everyday instruction or K+ (enrichment) programming. Children who were not identified at risk were place in alternating-day classes unless they attended a Title I School.

3. Generalization – Generalization of study findings will be limited to other small mid-western school districts with similar demographics.

A review of relevant literature is presented in the following chapter, which provides background information on effects of early childhood programs, definitions, and
review on the topic of quality and its relationship to student achievement in early childhood programs from child care through elementary school.
CHAPTER 2: LITERATURE REVIEW

Historical Background

When one considers the progression of American education, one element remains constant throughout history: change. The face of American education constantly evolves as the profession responds to trends, the needs of society, legislation, and new knowledge about how children learn (Fleischman, 2006; Plunkett, 1998; Reese, 2005; Weikart, 2000; Wolfe, 1998). As new information and ideas surface, educational practices change, often swinging back and forth between theories and practices like a pendulum, dependent upon the most widely promoted or most intriguing ideas and practices of the day rather than on solid research evidence regarding what constitutes effective educational practice (Plunkett, 1998; Slavin, 1989a; Slavin, 2003). Slavin (1989a) suggested that to make true generational progress, educators must stop this pendulum from swinging and redirect their efforts toward research-based strategies that help students achieve, rather than on ideas that “are merely new and sound good” (p. 752). Plunkett seemed to agree when she suggested that rather than simply stopping the pendulum from swinging, educators “must become increasingly sophisticated in accessing and applying this knowledge in our schools...while assuring that these applications are age-appropriate to the children with whom we work” (p. 313).

The need to make educational decisions based on sound research and empirical evidence that will result in long-term gains for students is particularly relevant in light of mandates associated with the No Child Left Behind Act. This federal legislation holds schools and districts accountable for providing quality education leading to academic
success for all students and all subgroups of students in public schools (U. S. Department of Education, 2004). Although, research on quality in education has gained momentum in recent years as growing demand from parents and state legislatures has raised the stakes for success in the classroom, the history of research on quality in education spans several decades (Katz, December 1999). Beginning in the 1960s, a number of researchers have focused on the effects of quality early childcare and early education. The researchers examined a number of variables, particularly the long-term benefits that result from quality programs (Barnett, 1998; Belsky et al., 2007; Marcon, 1996; Weikart, 2000).

The following pages include a definition of quality as related to early child care and education and present a brief overview of past and current research giving evidence to the lasting benefits of quality early child care and early childhood education in America. This is followed by a summary of research conducted to measure specific quality variables and the effects these have on student outcomes from early child care programs through elementary school. Finally, recent research is reviewed that addresses effects of global classroom and teacher quality on student outcomes, particularly as related to the development and use of the Classroom Assessment Scoring System as an evaluation tool (Pianta, La Paro, & Hamre, 2008). The discussion concludes with recommendations for how existing literature might be useful in providing a rationale for expanded research related to global classroom quality and how it impacts student outcomes.

Defining Quality

Clara Pratt, of the family policy program at Oregon State University, Corvallis, defined quality early childhood programs simply – those possessing “characteristics that
lead to positive outcomes for children” (cited in Galley, 2000, para. 7). However, providing an operational definition for quality is far from a simple endeavor. Quality in early child care and early education is viewed in a variety of ways. Some define quality subjectively, while others base their definitions on objective research. The approach to defining quality varies by the role and experience of the person offering the definition. But by any account, it is a “very difficult, complex and profound question” (Smith, 1996, p. 3). In fact, Smith argued that there is likely no completely objective definition of quality as researchers are inevitably “influenced by values and political context at every step of the process” despite efforts to remain objective (p. 7). The following definition of quality related to child care, attributed to Phillips and Howes (1987), was cited by Smith:

In research, quality has been viewed in several ways. First, global assessments of quality have been used to capture the overall climate of a program. Second, efforts to extract the specific dimensions of quality have emphasized (a) structural aspects of childcare, such as group composition and staff qualifications (b) dynamic aspects of child care that capture children’s daily experiences, and (c) contextual aspects of childcare, such as type of setting and staff stability. (p. 3)

Three basic factors were addressed in Phillips’ and Howes’ definition: structural, dynamic, and contextual variables (as cited in Smith, 1996). Smith asserted that while structural factors have been the focus of the majority of studies, perhaps because they are easier to assess, they do not guarantee program quality. Rather, she suggested that dynamic measures, also called process measures, are more likely the best indicators of quality. The researcher acknowledged the historical and practical value of past research conducted on structural aspects of early child care and education but directs the current
focus toward dynamic variables, particularly those over which teachers have power to implement change in their classrooms, while acknowledging the contextual variables present in the study environment.

Effects of Quality Early Childhood Programs

**Overall Benefits**

“Providing quality early childhood care and development programmes [*sic*] is a challenge for policy makers and educational planners all over the world” (Weikart, 2000, p. 7). Weikart suggested that for many years, education was thought to begin at the primary level and “preschool was considered a luxury available in those countries or for those families who could afford it” (p. 7). However, as roles and status of women have evolved and as women have become integrated into the paid labor force, there has been a growing understanding that providing quality early childhood education programs at all developmental stages must become a “national as well as a family concern” (Segal et al., 2006, p. 3; Weikart, 2000).

Providing high quality early childhood programs to improve student outcomes makes sense. According to Weikart (2000), “[t]here is a fundamental logic behind the belief that providing children with early stimulation and improved opportunities will create better performance later on as they tackle the demands of life” (p. 44). He cited a growing body of evidence from “carefully drawn studies of early childhood education programmes [*sic*] suggest[ing] a pattern of cause and effect that stretches from early childhood into the adult years” (p. 22). He stated that although a number of studies have examined the short-term effects of early childhood programs, a few “long-term and well-documented studies of the impact of early childhood care and education have reported
significant benefits for disadvantaged youngsters who have the opportunity to participate in high-quality programmes [sic]” (p.15).

In the following sections, significant studies and reports that address the issue of quality in early childhood programs are summarized. This is sometimes a confusing task as the histories of early child care and early education are intertwined. Many preschool-age children move from child care settings to preschool settings and sometimes back and forth between the two. In addition, while some studies discuss quality in the context of clearly defined developmental stages or clearly articulated program definitions, others address early childhood program quality in a more general sense or address quality at multiple levels within the same study. In spite of this, an attempt has been made to organize findings into two categories—research that primarily addresses overall effects of quality in early childhood and elementary environments and that which addresses specific quality variables in these settings. This discussion is followed by a brief review of research related specifically to quality at the kindergarten level. Finally, gaps in current research are identified and recommendations made for additional study focusing on the connection between program quality and student achievement in kindergarten.

Early childhood programs include those designed with a primary emphasis on custodial or physical care (child care) and those that have a more intentional emphasis on developmental goals for children (preschool and early elementary). Some might suggest that any quality program of care for children should attend to development of all domains: physical, social, emotional, and cognitive. For that reason, preschool quality is considered to include both child care and preschool environments in this study. Quality factors at the elementary level are addressed separately as indicated.
Quality early childhood programs are “increasingly recognized as being potentially beneficial for children regardless of whether they are living in poverty” (Segal et al., 2006, p. 3). This position was supported by Apple (2006) and Howes, Phillips, and Whitebook (1992) who posited that cognitive and social development benefits of participation in quality programs are well-documented. A number of additional studies indicate that quality child care is related to short-term and long-term benefits for children and society. One of these, conducted by Clark, Stewart, and Allhusen claimed:

One of the most robust findings in the early childhood literature is that good child-care quality is associated with a variety of positive outcomes for young children. Specifically, children in higher-quality child-care programs perform better on measures of social, language, and cognitive development when compared with other children. (Cited in NICHD, 2002, p. 199)

Yet another study, conducted by Peisner-Feinberg et al. (2001), provides additional “evidence that child-care quality has a modest long-term effect on children’s patterns of cognitive and socioemotional [sic] development at least through kindergarten” (p. 1534). The researchers claimed ongoing benefits of quality early childcare include longitudinal effects for language and math ability, “cognitive and attention skills, problem behaviors, and sociability, indicating that children who had better quality preschool experiences were more advanced in their development over a five-year period” (p. 1549). These findings were further supported by the National Institute of Child Health and Human Development [NICHD] Study of Early Care and Youth Development (2006), which reported higher quality care was related to “advanced cognitive, language, and pre-academic outcomes at every age and better socio-emotional and peer outcomes at some
ages” (p. 99). Additional support was offered by Fontaine, Torre, and Grafwallner (2006), who cited the Carolina Abecedarian study, which claimed children receiving high quality care demonstrate “optimal school readiness” and “higher cognitive test scores through age 21,” particularly in mathematics and reading (pp. 100-101).

“Policymakers at the federal, state, and local levels [also] recognize the key role preschool education plays in children’s learning and development” (Frede & Ackerman, 2007, p. 2). This is evidenced by past and current studies focusing on effects and quality of preschool programs (Barnett, 1995, 1998, 2004, 2008; Barnett, Hustedt, Friedman, Boyd, & Ainsworth, 2007; Barnett, Jung, Wong, Cook, & Lamy, 2007; Bryant, Clifford, Early, & Little, 2005a; Gorey, 2001; Schweinhart, 1994, 2005, Weikart, 2000). In a 2004 article, La Paro et al. claimed there was “near-record attention to providing access to high-quality preschool and early elementary educational programs” with state-wide prekindergarten programs being implemented in some areas and others “funding large-scale initiatives and pilot programs” (p. 409). These initiatives included both extensive and intensive studies examining overall effects of preschool education, specific quality variables, and how they affect children (Barnett et al., 2007; Schweinhart, 1994, 2005). The following review provides a brief sampling of studies in both categories; overall benefits and effects based on specific quality variables in early childhood education.

According to Mead (2008),

When it comes to pre-k programs, quality is the operative word. All of the research showing positive effects from pre-k focuses on programs that are of high quality. Lower quality programs do not achieve the same results, and research
suggests that extremely poor programs may actually be harmful to students.

(Mead, 2008, p. 25)

As stated previously, the overall benefit of quality preschool education for children is well-documented by a number of studies. One such study published by The National Institute for Early Education Research [NIEER], concluded “[h]igh quality preschool education produces substantial long-term educational, social, and economic benefits” (Barnett, 2004, p. 2). Another, conducted by the National Center for Early Development and Learning [NCEDL] supports this claim (Bryant et al., 2005a). Additional studies, such as the High/Scope Perry Preschool Project (Schweinhart, 2005), the Carolina Abecedarian Project (Campbell et al., 2002), and studies of the effect of Head Start (Devaney, Ellwood, & Love, 1997; Barnett, 2002) also support claims of long-term benefits for children attending quality preschool programs, particularly for children in poverty. In the absence of the quality indicator, however, such claims were not supported (Barnett, 2004). “In an era in which state-sponsored preschool programs are required to address academic standards and benchmarks, the extent to which variations on components of pre-K programs and aspects of classroom experience contribute to pre-academic skill growth is a critical focus for research” (Howes et al., 2008, p. 28).

One of the studies mentioned above, The High/Scope Perry Preschool Study, “is a scientific experiment that has defined both the short- and long-term effects of a high-quality preschool program for young children living in poverty” (Schweinhart, 2005, p. 1). In this longitudinal study, researchers identified 123 high-risk, low-income, African-American children between the ages of three and four to participate. About half of the
children were randomly assigned to a group that received high-quality preschool while the other half received no preschool program. These children were followed through age forty and data collected about them. The data included information such as: school dropout rates, academic achievement, employment, crime rates, and income levels. Study results support the positive benefits of quality programming on children's intellectual and social development with evidence showing educational, economic, social, and health benefits into mid-adulthood for the children who received preschool programming (Schweinhart, Montie, Xiang, Barnett, Belfield, & Nores, 2004; Weikart, 2000).

Another longitudinal study, the Carolina Abecedarian Project, begun in 1972, made similar claims about the benefits of quality early childhood programming for children at risk. This study followed children from infancy through young adulthood. Findings support claims that early intervention through quality early child care and education programs provide long-term academic and social benefits. Outcomes for the treatment group included higher intellectual and academic scores, a reduction in teen pregnancy, and attainment of more years of education and college than the control group (Bryant et al., 2005a; Campbell et al., 2002; Masse & Barnett, n.d.; Pungello, Campbell, & Miller-Johnson, 2000).

Research on a third early childhood program, Head Start, provides mixed support regarding its long-term benefits. Copple, Cline and Smith (1987) reported long-term benefits of the Philadelphia Head Start Program. They cited better attendance, fewer grade retentions, and fewer missed standardized tests among the positive effects of the program but concurred on prior analyses that showed little or no lasting effects on achievement scores for participants. Epstein (1995) also acknowledged long-term social
effects of Head Start including a reduction of students placed in special education and fewer grade retentions. However, he asserted that lack of significant long-term academic benefits overshadowed minimal social benefits. Lee and Loeb (1994) substantiated claims that academic benefits for Head Start children fade over time. However, rather than associating this with flaws in the program, they suggested that poor subsequent program quality undermined the positive academic benefits of Head Start and argued that social benefits alone are satisfactory reasons to continue the program.

Specific Quality Variables

Specific Quality Variables in Preschool Settings

As stated in the section defining early childhood quality, two types of indicators are generally used by researchers to identify and assess the quality of an early childhood education program.

One type is called “structural quality” - the characteristics of [a program] that can be regulated, such as level of teacher education and experience, number of children per class, and teacher to child ratio. The other way of looking at the quality of a program involves observing classrooms and the kinds of learning activities in the classroom, as well as the nature of the teacher-child interactions and children’s interactions with each other. Researchers call this “process quality.” (Bryant, Clifford, Early, & Little, 2005b, pp. 15-16)

Both of these types of quality in early childhood programs are addressed while emphasizing the necessity of an expansion of research on process quality in preschool and early elementary classrooms. Although research suggests that participation in quality preschool programs results in benefits for the children involved, early studies provide
limited information regarding which specific characteristics define quality programs. With so many factors associated with this determination, it is essential that researchers carefully examine program variables to establish which are related to positive benefits for children and which may result in less than ideal environments in which children develop and learn. Included among the research addressing specific factors contributing to program quality are studies related to structural variables such as effect of teacher qualifications and ratio of children per teacher and more recent studies that have begun to assess process quality, as well (Bryant et al., 2005b; Fukkink & Lont, 2007; Gerber, Whitebrook, & Weinstein, 2007; Peisner-Feinberg, et al., 2001; Zahorik, 1999).

In a 2002 study conducted by NICHD Early Child Care Research Network, researchers examined the inter-relationship of structural and process variables on children’s outcomes. They concluded that both types of variables affect children’s development directly and indirectly and therefore, must be considered when assessing program quality in the learning environment. Structural variables are typically simple to assess with objective measures, while process variables are more difficult to evaluate, due to their subjective nature (Belsky et al., 2007; Cassidy et al., 2005). At times, researchers address structural and process variables independently of one another. However, structural and process variables are sometimes viewed as inter-dependent and therefore, viewed in combination. When this is the case, it makes reporting, discussion, and analysis more complex than if variables could be addressed individually or in isolation.

*Child-staff ratios and class size.* Two important variables contributing to early childhood program quality are adult-child ratio and class size. The NICHD study (2002) reported that lower child-staff ratios predicted more positive child-caregiver interactions,
which have been associated with better developmental outcomes, such as increased language and skill levels for children in a number of studies. In addition, the same study reported improved social outcomes for children when group sizes were smaller. Textor (1998) shared similar findings in his report, in which he claimed when groups are too large, quality of interactions decreases and development is hindered.

Additional support for the importance of teacher-child ratios and class size in provision of quality preschool programs can be found in a number of additional studies, with results consistently showing smaller class sizes produce a range of better outcomes for children (Barnett et al., 2007; Denton, 2001; Mead, 2008; NICHD, 2002; Zahorik, 1999). According to Denton, low student-to-teacher ratios and small classes are among the five most important variables of preschool quality. Denton stated, “High quality preschool programs for 4-year-olds have no more than ten children per teacher and no more than 20 children in a class. For younger children, student-to-teacher ratios should be lower and classes smaller”, which facilitates more “individualized attention and nurturing interactions” (p. 21). These factors are associated with better overall quality and language development, as well. These claims were substantiated by the 2000 report by the National Research Council of the National Academy of Sciences, which made the claim that smaller classes and lower teacher-to-student ratios allow teachers to give children more individual attention and nurturing and are associated with higher global quality scores (Barnett et al., 2007).

Bryant et al. (2005a) reported an average of eight children per adult and an average group size of 18 children per class in the NCEDL Pre-kindergarten Study, which is within accreditation standards of the National Association for the Education of Young
Children (NAEYC). Findings from these studies show that when classes are within recommended ranges, programs tend to receive higher quality ratings than those that exceed recommended class-size limits.

**Physical environment.** Environmental factors, such as safety, cleanliness, equipment, and space also contribute to overall program quality in early childhood settings. Findings from a study conducted by the NICHD identify specific “regulatable” environmental characteristics of child care programs that affect quality. The researchers claimed environments that are safer, cleaner, and that provide more stimulation are “associated with better developmental outcomes for children”. In addition, they identified space per child as a structural variable affecting early program quality, with more space per child allowing increased opportunity for appropriate play and other forms of active learning (Peth-Pierce, 1998, p. 10). Others disagreed with the claim that more space is related to higher quality. Pessanha et al. (2007) found that space had little effect on process quality. Textor (1998) agreed. He cited studies that demonstrate that room size is not as important as long as over-crowding does not exist and the room is furnished with quality materials that address children’s needs and interests.

**Curriculum.** According to Mead (2008), quality prekindergarten programs must have a clearly articulated curriculum (p. 27). Curriculum, simply defined is “what to teach and how to teach it” (Frede & Ackerman, 2007, p. 2). This element adds a widely debated question regarding early childhood program quality and its effect on children-- does an academic curriculum in preschool provide challenges that enhance development, or impose pressure that can lead to future social and emotional problems for young children? According to Hirsh-Pasek, Hyson, and Rescorla (1990), results from their study
suggested “no academic advantages for children” exposed to highly academic environments. They warn of ramifications including “potential disadvantages in creative expression and emotional well-being” for children exposed to academic acceleration rather than developmentally appropriate curriculum (p. 401).

In an earlier article, Weikart (1981) discussed a group of studies begun in the late sixties to determine if curriculum differences based on varied theories of child development affect achievement and aptitude outcomes for children. After a meta-analysis of the findings, he concluded that “the basic issues in successful programming relate to quality of program implementation rather than philosophy of curriculum selected” (p. 33). Later, however, Weikart (2000) seemed to change his position when he commented on the necessity of “validated, well-implemented educational methodology” in the form of appropriate early childhood curriculum as a vital component of quality early childhood programs (p. 45). He argued that while it may seem logical to focus early childhood instruction on skills and content students need to know later in their schooling and life, this logic fails to account for the child as an “extraordinarily complex organism”, who “matures differently throughout childhood and adolescence” in a variety of domains (p. 44). Therefore, he posited early childhood curriculum should match the developmental needs of children at different stages rather than viewing children as smaller versions of adults with the same educational needs. The difference in these positions seems to be in the outcomes addressed. While early studies focused primarily on aptitude and achievement scores, later studies included a broader range of outcomes including cognitive, social and economic implications.
Schweinhart and Weikart (1998) and Weikart (2000) specifically cited the High/Scope Perry Preschool Project, begun in the 1960s, and the related High/Scope Preschool Curriculum Comparison Study as examples of controlled studies that have shown long-term intellectual and social benefits based on program curriculum. The High/Scope Perry Preschool Project is a longitudinal study that claimed benefits for participants in quality early childhood programs including: fewer behavioral problems; fewer repeated grades in school; lower rates of delinquency, decreased dependence on welfare and teen pregnancy; and increased rates of graduation and employment. The High/Scope Preschool Curriculum Comparison Study also provided evidence that curriculum model affects long-term student success, suggesting that child-initiated curriculum produces better long-term social and academic outcomes than direct instruction or eclectic models.

The D. C. Study (Marcon, 1996) also addressed the relationship between curriculum and student outcomes. In this study, three curriculum models (child-initiated, academically directed or middle-of-the-road) were compared to determine whether curriculum model affects student outcomes. Researchers concluded that “the type of preschool intervention [is] especially important, with the negative impact of didactic, academically directed preschool becoming most evident in the transition from third to fourth grade” when children are more likely to be required to engage in higher order thinking skills which involve creativity, independent thinking, and application of learning (p. 2).

A recent study conducted by the National Center for Educational Research [NCER], the Preschool Curriculum Evaluation Research Initiative [PCER], examined the impact of “14 preschool curricula on five student-level outcomes (reading, phonological
awareness, language, mathematics and behavior) and six classroom-level outcomes (classroom quality, teacher-child interaction and four types of instruction)” compared to control curricula (2008, p. xii). The study included 2,911 children, 315 classrooms and 208 preschools across the U. S. Early findings suggested that curriculum differences affect both student- and classroom-level outcomes, but impact varies by program, with no program demonstrating significant effects in all areas. A weakness of this study seems to be in the lack of standardization of program controls used with non-treatment classes, limiting claims of program effectiveness. While the study addressed a number of curriculum models and provided extensive data about these, it did not provide a means of comparison between programs. Thus, while the PCER study provided a starting point for future research, it did not provide adequate data to make determinations regarding which curricula produce better quality preschool programs.

The controversy over what constitutes appropriate curriculum for early learners persists. According to Armstrong (2007), “A [curriculum] superhighway is being built across today’s education landscape” that links academic standards from early childhood through early adulthood, pushing children around the academic racetrack at breakneck speed rather than focusing on the whole child with its varied and complex set of needs (p. 16). Rothstein, Wilder, and Jacobsen (2007) provided a similar warning, that public education should avoid a purely academic curriculum, but should instead provide balance that goes beyond basic skills and test scores to “produce the outcomes necessary for success in work and life” (pp. 8-9).

According to Tomlinson and Germundson (2007), even exceptional programming is ineffective if it’s all the teacher has to offer. Quality education goes beyond
curriculum; it takes excellent teachers to make the most of the resources at their disposal. They argued that great teachers are needed to create the balance necessary to make “music” of educational practice. Perhaps they were right; quality curriculum is not sufficient to insure quality education. Perhaps, quality teachers are also essential elements in the provision of high quality learning experiences for all children. This leads to the next variable that affects early childhood program quality.

Teacher quality. A number of past studies examined teacher or caregiver quality as an indicator of quality in early childhood programs. Teacher quality has been equated with the level of training attained by caregivers (Textor, 1998). In a meta-analytic review of child care quality studies published between 1980 and 2005, Fukkink and Lont (2007) concluded that specialized training for caregivers showed significant positive results for the caregiver and for the children in their care. They stated, “[T]he current empirical evidence demonstrates that specialized training improves the pedagogical competencies of caregivers in childcare, including their professional attitude, knowledge, and skills” (p. 305). Pessanha, Aguiar, and Bairrão (2007), and others added to this claim. Pessanha et al. suggested that younger and better-paid teachers provide higher quality care for toddlers, demonstrating the inter-connection between a structural variable (teacher salary) and a process variable (care), both of which affect outcomes for children.

If the statements above are true, they lead to an important question in early child care and education, the degree to which student outcomes can be improved by elevating teacher education requirements (Kelly & Camilli, 2007). Perhaps Barnett (2004) was on the right track when he proposed that benefits of early childhood programs are large “when teachers are professionally prepared and adequately compensated” (p.2). He
stated, “New research indicates that young children’s learning and development depend on the educational qualifications of their teachers. The most effective preschool teachers have at least a four-year college degree and specialized training in early childhood” (p. 1). But, perhaps this conclusion does not go far enough or include all critical variables which contribute to teacher quality.

According to Gerber et al. (2007), more teacher training results in higher program quality and greater teacher sensitivity toward children. However, another study, conducted by Torquati, Riakes, and Huddleston-Casas (2007) provided mixed results regarding the effects of teacher preparation on quality of care. Their study showed that holding a Child Development Associate (CDA) certificate, more years of education, and child-development coursework predict global observed quality when working with toddlers, but not necessarily when working with infants.

In addition to the above findings, other studies showed null or contradictory associations between level of teacher education and program quality. In a meta-analysis of seven major studies of early care and education, Early et al. (2007) found that policies which focus only on teacher education are insufficient in determining the effects of teacher quality on student outcomes. They suggested that although teacher education matters, teacher quality goes beyond credentials; raising the effectiveness of early childhood educators will require a range of supports including professional development and specialized training to improve teacher-student interactions.

Others agree that measures of teacher quality should extend to teacher effectiveness and the quality of teacher-student interactions in addition to minimum education or credential requirements. Perhaps as these researchers propose, teachers
should be accountable for the outcomes they produce as well as the quality of interactions they have with children (Commission, 2007; Darling-Hammond & Berry, 2006). This rationale was supported by Howes et al. (2008) when they cited Bowman et al. and Pianta with the following statement:

Effective teaching includes sensitive interactions with adults around instructional content within a positive social and emotional classroom climate and specific instructional content. Children learn in the context of interactions with adults; this seems to be particularly the case for young children’s learning of pre-academic skills related to early literacy, language development, and task-orientation. (p. 29)

A number of large-scale studies have examined teacher effectiveness, particularly how teacher-child relationships affect student outcomes and how teachers establish global classroom quality through creation of a positive environment, strong classroom management, and quality feedback (Howes et al., 2008; La Paro et al., 2004). Two such studies, the NCEDL Multi-State Study of Pre-Kindergarten and the State-Wide Early Education Program Study [SWEEP], provide relevant data from pre-kindergarten programs across eleven states. Results from these studies indicate that “gains in language-related academic skills are greater largely as a function of classroom processes directly experienced by children, particularly the instructional climate of the classroom and the teacher-child relationship quality,” elements which are controlled or at least affected by the teacher (Howes et al., 2008, p 45).

*Length of day.* The amount of time children spend in an early child care or preschool environment provides one more quality consideration in early childhood programs. Burchinal suggested that evidence from studies showing benefits from full-day
over half-day kindergarten may be applicable in preschool programs as well (cited in Bryant et al., 2005b). According to Zill et al. and Walston and West, “Full-school-day participation in [pre-kindergarten] and kindergarten has been found to produce larger cognitive gains than half-day participation” (cited in Bogard & Takanishi, 2005, p. 13). They also concluded that full-day programs better prepare children for elementary school than half-day programs. Additional research suggests that program length affects the benefits afforded by preschool programs. While evidence suggests that full-day programs benefit children, particularly those coming from disadvantaged backgrounds, studies also suggest that too much time spent in preschool may have negative effects, particularly on children’s behavior (Peth-Pierce, 1998; Love et al., 2003).

**Summary of Quality in Child Care and Preschool Settings**

A number of researchers concluded that the short- and long-term psychological and academic benefits of quality early childhood programs are positive, especially for children from disadvantaged backgrounds (Weikart, 2000). Quality preschool programs are characterized by researchers as those possessing essential indicators of structural quality including: safe environments, highly qualified teachers, provision of appropriate equipment and resources, developmentally appropriate curriculum, small classes, low teacher to student ratios, and appropriate length of school day. Process variables further characterize program quality in the form of positive relationships between teachers and children, high levels of communication between teachers and children and teachers and parents, positive emotional climate, instructional support, and strong classroom management (NIEER, 2002; Pianta et al., 2008). It therefore makes sense to invest in further research designed to expand understanding of quality variables related to these
positive outcomes and to determine if these quality variables and their effects hold true at the elementary level.

**Quality Variables in Elementary School**

**Overall Benefits of Elementary School**

While studies have been conducted related to a number of quality variables at the elementary level, large bodies of research address two specific quality variables at this level—class size and teacher quality. While prior discussion addresses these variables in the context of preschool, the following sections address these variables and their related outcomes specifically related to formal schooling in the elementary setting.

**Specific Quality Variables in Elementary Settings**

**Class size.** Several major studies provide evidence that class size affects outcomes for children at the elementary level (Blatchford, Bassett, Goldstein, & Martin, 2003; Smith, Molnar, & Zadorik, 2003; Nye, Hedges, & Konstantopoulos, 2001; Pate-Bain, Boyd-Zaharias, Cain, Word, & Binkley, 1997), with a number of additional studies providing support for the positive effects of small class size (Achilles et al., 1997/1998; Ellis, 1984; Finn, Gerber, Achilles, & Boyd-Zaharias, 2001; HumSanito, Hunntosi, & Tessling, 2001; NICHD, 2004; Yan & Lin, 2005). Blatchford et al. claimed results from their study indicate clear effects of class size difference on children’s academic achievement over the first year. They specifically noted differences in literacy achievement and highlighted evidence that students with the least skill at entry were most helped by smaller class size, especially when the class size was smaller than 25 students.

Multiple related effects of smaller classes were also shown to contribute to processes used in large and small classes. More difficulties were found in larger classes
where there were more large groups, including more off-task behavior and less attentiveness. Smaller classes allowed for more individual contact with the teacher and more support for learning. These findings were supported by Hunn-Sanito et al. (2001), who suggested smaller classes make teacher workloads more manageable, allow more time for work on basic skills and teacher feedback, and allow teachers to pinpoint students who need extra help, resulting in fewer grade retentions and referrals for special education.

Results from a five-year longitudinal study conducted by Smith et al. (2003) yielded similar findings to those stated above. The researchers asserted the Wisconsin Student Achievement Guarantee in Education Study (SAGES) shows a clear effect for class size, particularly for low-income students. They claimed the largest gains were seen in first grade, with a 25-30% increase in achievement for children in smaller classes over their counterparts in larger classes, with effects persisting into third grade. They also indicated that although class size reduction benefits all children, benefits for African American students were particularly strong.

The Student/Teacher Achievement Ratio (STAR), another significant longitudinal class-size study, provided additional convincing evidence that small classes impact academic achievement immediately and positively. The Lasting Benefits Study and STAR Follow-up Studies: 1996-1997, provided follow-up of Project STAR by tracking students through grades seven and ten. These reports reinforced findings from the original study, showing continued benefits in the form of higher achievement in mathematics and reading, along with reduced number of grade retentions and drop outs. Researchers in both studies concluded that effects continue for years after children return
to regular-sized classes and are large enough to promote educational policy supporting smaller classes, particularly for minority students and those with free lunch status who tend to benefit the most from class-size reduction (Hunn-Sanito et al., 2001; Nye et al., 2001; Pate-Bain et al., 1997).

Teacher quality. “Studies show that well-prepared and well-supported teachers are important for all students…that is why one of the most important aspects of NCLB is its demand that states ensure a ‘highly qualified’ teacher for every student” (Darling-Hammond & Berry, 2006, p. 15). NCLB, which has been identified as a driving force behind education reform, places unprecedented emphasis on teacher quality at all levels, requiring states to ensure that all teachers are “highly qualified”. This is defined as earning a minimum of a bachelor’s degree, holding state licensure or certification, and demonstrated knowledge in the subjects taught (Commission, 2006; U.S. Dept. of Education, 2004). The Commission on No Child Left Behind (2007) provides agreement in a report on the progress of the law. It states, “One of the foundational principles of NCLB, supported by ample research, is the idea that teacher quality is the single most important school factor in student success” (p. 30).

While the writers of the No Child Left Behind Act and a number of researchers suggest teacher qualifications are an indicator of quality, others suggest that qualifications alone do not insure effectiveness (Carey, 2004). Rather, it is suggested that teacher quality should “focus on teacher effectiveness in improving student achievement rather than qualifications for entering the profession” (Thompson & Barnes, 2007, pp. 1-2).
“Almost nobody, it seems, disputes the importance of effective teachers, including teachers themselves” (Haycock, 2004, p. 1). Carey (2004) supported this position with findings from analyses of studies conducted in Tennessee and Texas. He reported findings in Tennessee, which claimed, “all else being equal, students assigned to the most effective teachers for three years in a row performed 50 percentile points higher…than comparable students assigned to the least effective teachers for three years in a row” (p. 4). Carey also suggested the impact of teacher effectiveness “exceed[s] any one thing about the students themselves” identifying teacher effectiveness as “the single biggest factor influencing gains in achievement,” with greater effects than any other factor associated with student success or failure (p. 4). The Texas study substantiated the Tennessee findings, showing “teacher effectiveness varied dramatically and had a major impact on student performance,” an effect substantial enough to “offset or even eliminate the disadvantage of low socio-economic background” (p. 5).

Although NCLB affirms a child’s right to a highly qualified teacher, since the legislation became law in 2002, states have scrambled to define and regulate exactly what “highly-qualified” means. What constitutes teacher quality at the elementary level? According to Traina (1999), a historical review of the characteristics of good teachers found in autobiographies of prominent Americans reveals a consistent trio of descriptors: competence in subject matter, caring for students, and distinctive character. While few would argue with these attributes, Pianta (2007) suggested that teacher quality is more commonly defined by a number of proxies, described in terms of: level of education, number of years of experience, certification, and “consistent production of test-score gains among their students” (para. 3). However, he argued, there is a weak link between
these proxies and students’ academic progress. He suggested alternate measures of teacher quality, which should be based on a combination of variables including: assessment of students, interactions in the classroom, and use of higher-order thinking skills.

An early review of research conducted by Brophy (1986) summarized research that links teacher behavior to student outcomes. Findings suggested that teacher quality is related to teachers’ perceived roles, their expectations for learners, and the amount of time they allocate to academic activities, as well as their ability to manage the classroom and keep students engaged. In addition, he suggested that students achieve more in classes in which teachers are actively engaged in teaching and involved interactively with students rather than in passive supervision.

Another study, conducted by Rubie-Davies (2007), identified an additional factor in teacher quality: teacher expectations. She contrasted outcomes for teachers who hold low, average, and high expectations for their students. Results show that average and high-expectation teachers spend more time scaffolding, clarifying, questioning and providing feedback to students than teachers with low expectations. Differences in expectations result in differences in socio-emotional and instructional environments. According to Rubie-Davies, these effective strategies, which are based in teacher-student interaction, also tend to result in substantial differences in reading progress for children.

Darling-Hammond and Baratz-Snowden (2007) also connected teacher-student interactions to a description of quality teachers. They identified high levels of student engagement, clearly articulated expectations, ongoing feedback, and the ability to design a “well-functioning, respectful classroom that allows students to work productively” as
essential features of high-quality teaching (p. 112). In addition, they proposed that high-
quality teachers collaborate effectively with parents, other teachers, and administrators to
reduce obstacles and create a supportive environment for learning.

Numerous authors identify characteristics of quality teachers; however, a
relatively new trend in educational research goes beyond describing good teachers to
measuring the effects teacher quality have on student outcomes, with several recent
studies focusing on the role of teacher-child relationships and student success as part of
the equation defining teacher quality. Quite a few of these studies involved Robert Pianta
and his colleagues at the University of Virginia, Charlottesville Center for Advanced
Study of Teaching and Learning [CASTL]. The center’s mission is to “improve
educational outcomes through the empirical study of teaching, teacher quality, and
classroom experience from preschool through high school, with particular emphasis on
the challenges posed by poverty, social or cultural isolation, or lack of community
resources” (http://www.virginia.edu/vpr/CASTL/). Studies conducted by CASTL
researchers and collaborators examined the relationship between teacher-child
relationships and children’s success in school. Findings suggest that relationships
between teachers and their students have a substantial impact on social and academic
development for children. Further, these studies suggest that emotional and instructional
support offered in the context of positive teacher-child relationships has the potential to
mediate other risk factors for children and affect children’s ability to attain the
competencies necessary for school success (Hamre & Pianta, 2001, 2005;
La Paro, Pianta & Stuhlman, 2004; Pianta, Stuhlman, & Hamre, 2002; Saft & Pianta,
2001; Stuhlman & Pianta, 2001).
Summary of Quality in Elementary Settings

Substantial literature provides evidence that a number of variables contribute to quality in early child care, preschool and elementary programs and that the effects of quality in each setting can provide positive benefits for the children involved. However, because of the differences between preschool and elementary program designs and goals and those of kindergarten programs, these findings may or may not hold true at the kindergarten level. The following section discusses studies examining kindergarten quality variables and effects and recommends additional research to address gaps in existing literature that may provide beneficial information for improving schooling and outcomes for children.

After quality is defined and observed, whether it is related to structural or process variables, classroom or teacher characteristics, “its correlates and consequences can be examined systematically” (Pianta et al., 2002, 227). As stated previously, a number of large-scale studies provide direction for examining quality and related effects at the early childhood and elementary levels. The next logical step may be to apply this process at the kindergarten level, which has received “surprisingly little” attention to date, considering the “long record of similar research in child care [as well as prekindergarten and elementary] settings” (p. 226).

Quality in Kindergarten

Overall Benefits of Kindergarten

“Kindergarten is a critical period in children's early school careers. It sets them on a path that influences their subsequent learning and school achievement. For most children, kindergarten represents the first step in a journey through the world of formal
schooling” (U. S. Department of Education, 2000, p. v). While kindergarten is now almost universally available to children across America, with more than 93% of all five year-olds attending kindergarten in some form (Ackerman, Barnett, & Robin, 2005), quality research studies seem to be limited when it comes to addressing this critical stage of early education. Many early kindergarten programs in America were intended to ease the “acculturation of newly arrived immigrant children” to our country (Moyer, 2001, p. 82), but contemporary programs are often designed as the entry point into formal academic education. Unlike early child care programs, which tend to emphasize physical care and state-funded preschool programs, initially developed as compensatory programs primarily for children at risk due to economic or other factors related to learning delays, early American kindergartens were established with an emphasis on academic development in young children (Segal et al., 2006). The contemporary purpose of kindergarten, while viewed as different from that of child care and preschool programs, has also been considered separate from and different than that of primary/elementary schools. In recent years, programs originally designed as child-centered, developmental, transitional, and/or readiness programs have become more and more academic in nature. Accountability and increased expectations at the primary level have created a push-down effect with more attention directed at the academic effects of kindergarten programs previously viewed as a transition between the preschool years and structured academic schooling of children at the elementary level (Goldstein, 2007; Katz, December 1999).

According to Pianta et al. (2002) increased interest in the quality of environments in public kindergarten classes is propelled by evidence of the importance of early education for later school success, with research supporting claims that attendance in
kindergarten, the type of instruction that occurs, and the kind of interactions that occur impact children’s achievement. However, much of the research at the kindergarten level focuses on structural variables rather than process variables, such as those identified by Pianta. Among the other quality variables studied, many overlap those at the preschool and elementary levels providing little new or different information.

The following sections discuss three areas of quality research that specifically address issues at the kindergarten level. Extensive research examines class-size, curriculum, and length of day. While these are also discussed at the preschool and elementary levels, research examining these variables specifically at the kindergarten level is significant and should be considered when determining kindergarten program quality and its effects.

Specific Quality Variables in Kindergarten

Class size. Similar to findings at the preschool and elementary levels, studies show that class-size affects behavioral and academic outcomes for children at the kindergarten level. One particular study, conducted by Finn and Pannozzo (2004) examined behavioral effects of class size for kindergartners. The researchers concluded that overall class behavior is better in kindergarten classes of twenty students or less than in larger classes. Achilles, Kiser-Kling, Aust, and Owen (1995) and Johnston (1990) suggested that behavior differences, at least in part, may be related to an increase of individualized instruction and smoother transitions that occur in smaller classes. Johnston stated, “Teachers in small size classes were found to use more desirable classroom practices such as more attention to individual children, and more individualization of
instruction” (p. 2), which may in turn translate into increases in academic performance for these students.

Additional studies examined academic effects of class size for kindergartners. In an early study conducted by the Chicago Board of Education, researchers concluded that smaller classes rather than the length of school day had the greatest impact on student achievement (Bridgman, 1986). More recently, a study conducted by Yan and Lin (2005) using data from the Early Childhood Longitudinal Study-Kindergarten Year found that smaller class sizes yielded a slight positive relationship with children’s mathematics and reading achievement, with most significant differences found for children coming from lower-socioeconomic and minority backgrounds. These findings were substantiated by yet another study conducted by Haenn (2002) in the Durham, North Carolina, public school system. This study showed higher test gain scores for kindergarten children in smaller classes compared to the control group.

In addition to findings regarding effect of class size at the elementary level, The Tennessee Student Teacher Achievement Ratio (STAR) study and follow-through studies of the STAR Experiment show specific benefits at the kindergarten level that are sustained over time. Pate-Bane and Jacobs (1990) reported “small class size is therefore a significant factor in kindergarten reading readiness and achievement” (p. 3) with follow-up studies “indicating that small classes lead to significant improvements in reading and mathematics and benefits are greatest for students who start in small classes early (full-day kindergarten or first grade)” (Resnick, 2003, p. 1).

Curriculum. Curriculum issues at the kindergarten level are similar to those at the preschool level in spite of differences in the purposes of the programs. According to Katz
(December 1999), “disputes concerning curriculum and teaching methods go back a long way in the field of early childhood education” (para. 1). She suggested a number of variables may account for increasing pressure to move away from developmental programs to those that introduce children to academics at earlier ages. Among these is the expectation that kindergarten programs should ensure readiness for higher grades. In addition, there has been a decrease in the value of spontaneous play as a natural mechanism for learning. Thus, the debate between “‘instructivists’ and ‘constructivists’ revolves around the extent to which formal academic instruction may be appropriate or even essential” for children whose early environments may not provide adequate opportunities for informal learning to take place (para. 6). Katz (2007) suggested that curriculum should focus on standards of experience rather than meeting academic expectations that overlook “the centrality of understanding as an educational goal” (p. 95).

Others, such as Goldstein (2007) and Harrington-Lueker (2000), suggested that at least part of the debate about curriculum can be attributed to the emphasis on standardized assessments which do not take into account the developmental differences that naturally occur within same-age groups of children. As teachers are held accountable for teaching standardized content and for outcomes on high-stakes tests, many feel torn between their philosophies of best practice for early learners and state mandates requiring standardized testing at the primary level. Harrington-Lueker reported that “while such high stakes tests are typically delayed until third or fourth grade, schools nationwide report increasing pressure” to prepare younger children “for testing and even to change their curricula” (para. 6) to match assessments.
“When using traditional measuring sticks of achievement tests and report card grades, it is difficult to say whether child-directed or didactic programs are superior” (Dunn & Kontos, 1997, para 7), however, Dunn and Kontos reported findings that suggest higher scores on measures of divergent thinking and creativity, as well as better language and increased confidence levels for children who attend child-initiated programs. Others disagree, claiming that teacher-directed, academically focused programs are necessary to provide appropriate education for young children, particularly those who come from less enriched backgrounds (Nelson & Rogers, 2003; Stone, 1996). Disagreement continues as some insist quality is defined by factors such as the child’s or teacher’s role in initiating learning, while others look only at outcomes on standardized assessments of academic skills as indicators of quality kindergarten curriculum. Perhaps, as Hyson (2003) and Egertson (2004) suggested, this is not an either-or dilemma, but one in which balance and individually determined needs should mandate what constitutes quality programming for kindergarten children.

**Length of day.** Using data from studies showing the value of early education in long-term academic achievement and evidence from schools that have offered it, more and more states are recognizing full-day kindergarten as the answer to a number of problems in education (Wood, January 29, 2004). This statement represents a major trend in contemporary kindergarten programs that has occurred over the last two decades, an expansion in the occurrence of all-day kindergarten compared to part-time programs (Walston & West, 2004). Using statistics from The Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999, researchers at the National Center of Education Statistics [NCES] compared effects of full-day versus half-day programs. Findings
suggested that while program structure is similar between full- and half-day programs, full-day kindergarten is related to greater gains in mathematics and reading/language arts even “after adjusting for gain score differences associated with race/ethnicity, poverty status, fall achievement level, sex, class size, amount of time for subject area instruction, and the presence of an instructional aide” (p. xxi). These differences hold true for children across socio-demographic backgrounds and other characteristics of the classroom, as well.

A number of additional studies focused on the impact of length of day on kindergarten outcomes, all of which supported findings of the national study including increased academic growth in language arts and mathematics (Baskett, Bryant, White, & Rhoads, 2005; Clark & Kirk, 2000; Gullo, 2000; Lee, Burkam, Ready, Honingman, & Meisels, 2006; Nunnelly, 1996; Plucker et al., 2004; Rothenberg, 1995; Zvoch, Reynolds, & Parker, 2008). According to Plucker et al., “When analyzed on the major dimensions of academic achievement, grade level retention, special education referrals, and social and behavioral effects, the benefits of full-day kindergarten are apparent” (p. vi). According to Miller (2003), “No studies to date show greater gains, academic or developmental, for students in half-day programs over those for students in full-day programs” but do suggest long-term learning increases are related to full-day programs (p. 2).

Summary of Quality in Kindergarten Settings

Although the benefits of full-day kindergarten seem to be clear and overwhelming, it is not offered in all states for all children, and most teachers do not have a choice as to whether they provide full-day or half-day programming for their students.
Indeed, while each of the factors discussed above may have an impact on learner outcomes at the kindergarten level, these are elements over which the teacher has little control. The teacher cannot determine class size, length of day, the ratio of adults to students, or even the curriculum imposed in many instances. However, there are a number of variables over which teachers do maintain control in their classrooms. These relate to their roles in the classroom: as they establish and maintain an appropriate climate for learning; respond to students’ needs with sensitivity; manage behavior, resources, and the environment; and provide instructional support and quality feedback.

One might note that although classroom processes have been reported to have an impact on learner outcomes, the majority of quality research continues to focus on structural elements, particularly at the kindergarten level. While structural elements have been shown to impact program quality, these are variables over which teachers typically yield little control. In addition, the vast bulk of quality studies seem to focus on the preschool level, prior to formal entry to academic education or the elementary level, after children have already been part of the system for a year or more. There seems to be a clear gap in research addressing process quality at the kindergarten level, particularly examining the relationship between process quality and student achievement.

As La Paro et al. (2004) suggested, a number of factors should direct us toward increased attention to process quality in kindergarten classrooms. Among these is considerable literature that supports a relationship between children’s classroom experiences and their immediate and long-term social and academic development. In addition, there is an “unprecedented level of investment in and attention to early education” and “a policy climate that emphasizes accountability” (Pianta et al., 2008, p.
6). While accountability assessments by and large focus on direct measures of student achievement, this method alone is limited. Evidence indicates that direct assessment of early learners is less valid and less reliable than it is for older students due to a number of factors. In addition, if learner competencies are related to classroom experiences, it simply seems logical to assess the quality of those experiences as part of the accountability equation.

In 2004, La Paro et al. stated that there was a need to examine findings from the CLASS in relation to child outcomes. In response, a number of studies using the CLASS have been conducted at the preschool and elementary levels, which have consistently shown relations between classroom-level indicators and engagement in learning, as well as student achievement outcomes (Downer, Rimm-Kaufman, & Pianta, 2007). At the kindergarten level, studies utilizing the CLASS have also demonstrated a relationship between teacher-child relationships and learner outcomes, but these seem to focus on a connection between classroom interactions and student behavior rather than examining academic achievement (Pianta et al., 2002; Rimm-Kaufman, La Paro, Downer, & Pianta, 2005). Researchers have begun to investigate the association between classroom process quality and outcomes. Nevertheless, part of the picture is still missing—an appraisal of the effect of classroom process quality and student-teacher interactions on achievement at the kindergarten level. This study expands upon the current research by addressing this gap. Dynamic dimensions of classroom quality are identified, but the examination goes beyond identification to seek and identify connections between these indicators and student achievement.
After quality is defined and observed, whether it is related to structural or process variables, classroom or teacher characteristics, “its correlates and consequences can be examined systematically” (Pianta et al., 2002, p. 227). As stated previously, a number of large-scale studies utilizing the CLASS provide direction for examining quality and related effects at the early childhood and elementary levels. The next logical step may be to apply this process at the kindergarten level, which has received “surprisingly little” attention to date, considering the “long record of similar research in child care, prekindergarten, and elementary” settings” (p. 226).

Classroom Assessment Scoring System (CLASS)

While an extensive body of literature provides support for the claim that early childhood program quality affects children’s development, a number of recent large-scale studies examined more specifically how children’s experiences in school are related to social and academic outcomes. Many of these studies utilized the CLASS as a means of measuring classroom quality. Findings from these studies reinforce the notion that perhaps classroom processes are more relevant than structural program features for student development (La Paro et al., 2004). One such study, which utilized data collected from across eleven states, shows that most preschool classrooms offer moderately high emotional support, but relatively low instructional support, particularly “with regard to concept development and feedback, which were in the low or low moderate range” (p. 420). Another study shows that children attain larger gains in academic outcomes when closer teacher-child relationships and higher quality instruction are experienced (Howes et al., 2008). If these findings can be replicated, perhaps this tool will become a valuable
instrument for assessing quality features and improving quality in classrooms, consequently increasing achievement.

Use of the CLASS has become somewhat widespread. However, this investigation provides a unique application of the instrument by providing insight into the connection between classroom process quality and student achievement at a critical point in children’s education; their entry into the formal education system at kindergarten. As accountability and expectations for academic performance have increased, more attention has been directed at kindergarten programs, with the expectation that children at this stage should develop skills and behaviors that will help them succeed in later grades. Thus, examining classroom quality, children’s experience within classrooms, and related outcomes may provide a valuable mechanism for identifying and improving quality variables and ultimately, improving student achievement while addressing a need identified by developers of the instrument. La Paro, Pianta and Stuhlman (2004) asserted, “[F]indings from the CLASS need to be examined in relation to child outcomes” (p. 423).

This study is designed to directly address this need by providing a systematic and programmatic focus on classroom quality at the kindergarten level, which is especially important for “young children, whose developing skills are deeply embedded in their interactions with teachers and one another” (Pianta & La Paro, 2003, p. 28).
CHAPTER 3: METHODOLOGY

Purpose

This quantitative study focuses on the relationship between classroom quality and children’s academic achievement. Specifically, it examines how classroom quality in three broad domains—emotional climate, classroom management and instructional support—impact kindergarten achievement growth in mathematics and reading. The following questions were posed:

Research Question 1: Is kindergarten achievement growth in mathematics related to global classroom quality as measured by the CLASS?

Research Question 2: Is kindergarten achievement growth in reading related to global classroom quality as measured by the CLASS?

Research Question 3: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in mathematics?

Research Question 4: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in reading?

To answer the above questions, the researcher used mixed linear model analysis (hierarchical linear model) to compare classroom quality scores to student achievement in mathematics and reading.

Setting

The setting for this study was a small school district situated in a rural agricultural community near a large city in the Midwest. This district was selected because it is typical of many small school districts across the Midwest and because the researcher was
able to gain entry into the site through professional contacts with district administrators and prior collaborations with school personnel. The district served approximately 6,100 students in seven elementary schools, two middle schools, one elementary/middle school, and one high school during the study.

Schools and Classes

School populations varied across the district, with enrollment ranging in elementary schools from approximately 250 at the smallest school to over 450 at the largest. Of the eight elementary buildings in the district, two were designated Title I schools with two more schools serving additional students through partial Title I funding. Due to funding formulas, there was variety in programming provided for kindergarten children in the district. There were twelve all-day every day classes serving 200 children with the remaining students (approximately 250) served by alternating-day programs or extended day programs. This information was included in the coded data for each child and class so that program model was accounted for in the analysis (Corporation Snapshot, 2008; Enrollment Summary, 2008).

Participants

Teachers

The participants for this study included 24 kindergarten teachers who taught 28 classes of children in eight elementary buildings in the district. Of these teachers, all were Caucasian, 23 were female, and one was male. While all teachers in the district held an Indiana teaching license, level of education and experience varied. Seventeen teachers held a bachelor’s degree, and seven teachers held a master’s degree. Two of the teachers holding master’s degrees also had earned credits beyond the master’s level. Years of
experience also varied among district kindergarten teachers, ranging from only one year of classroom experience to over 27 years for the most experienced, with an average of approximately 13 years of teaching experience for kindergarten teachers in the district.

All kindergarten teachers in the district willingly participated in the study. Although the district superintendent and assistant superintendent of curriculum and instruction were supportive of the research project, teacher participation was not required, and teachers would not have been penalized if they had chosen to opt out of the study. To encourage all teachers to consent to observation and inclusion, the researcher offered to share aggregated outcomes and final study findings with participants and assured teachers that all data included in the study would remain anonymous for the purposes of publication.

Demographic information for teacher participants and classes was provided by the school corporation and direct submission by the teachers through a researcher-created survey (see Appendix A). In instances in which data were replicated, the researcher checked for consistency and contacted the teachers and/or corporation offices directly to verify missing or conflicting information. The following data were collected by means of direct teacher survey: teacher name, school, program (all-day every-day, alternating day, or K+), teacher’s degree (bachelor’s, bachelor’s plus additional coursework, master’s, master’s plus additional coursework, specialist, or other), number of years teaching (including current), number of years teaching kindergarten, number of students, number of assistants (and hours available), number of other adults in the classroom, number of students with special needs, numbers of children receiving speech services, and any special circumstances of which the teacher chose to inform the researcher. Teacher and
class data provided by the district as part of the NWEA score report included: special education designation (including high ability) for all students, meal pay status for students, LEP (Limited English Proficiency) designation, program schedule (Monday, Wednesday, Friday or Tuesday, Thursday, Friday), school, and Title One designation.

Students

Data for all kindergarten students ($n = 439$) who attended school in the corporation during the school year were provided to the researcher in spreadsheet format. However, only data for kindergarten students who were tested during all three testing sessions (fall, winter, and spring) were included for analysis. One additional student was excluded from the study because he/she was exempt from testing due to special needs identification and his/her Individual Education Plan.

The corporation provided data for kindergarten students including: identification number; gender; ethnicity; free, reduced or full-pay meal status; Title I designation; special needs designation (high ability or other exceptionality); limited English proficiency; school and teacher of record; and NWEA scores. These data were coded and archived by the school district and provided to the researcher in the form of PowerSchool© reports (Pearson, 2006). The researcher received only data sets for students. Names were withheld to insure confidentiality of personal information.

Data for all students in the district were included in analysis, with the only exceptions being those students who entered the district after initial assessment, left the district or changed teachers prior to post-testing ($n = 34$), or children who were exempt from testing due to identified special needs ($n = 1$). The final number of students included
in the analysis was 404. In summary, fewer than ten percent of students moved during the school year or were not present for all three testing cycles.

The student population in the district was comprised of approximately 95% white students and five percent non-white students, which was consistent for kindergarten children in the district. Of the almost 450 kindergarten children served by the corporation’s schools, 25 were non-white, of which two were American Indian, two African American, four Pacific Islander, four Hispanic, and thirteen multiracial children in the kindergarten program. Gender of children in the district was almost evenly distributed, with a difference of just over 100 more males than females district-wide and a difference of only two more males than females at the kindergarten level (Enrollment Summary, 2008).

Approximately 32% of the district population received free or reduced meals and approximately 17% of total school enrollment was identified for special education services. At the kindergarten level, however, the number of free and reduced meals increased; approximately 45% percent of the children received subsidized meals. Of these, approximately 35% received free meals and approximately 10% received reduced cost meals (Enrollment Summary, 2008).

Measures

Two measures were used to assess classroom quality and student achievement. The CLASS (Classroom Assessment Scoring System) was used to measure three aspects of global classroom quality (Emotional Support, Classroom Organization, and Instructional Support), as well as specific quality indicators, and The Northwest
Evaluation Association MAPS (Measures of Academic Progress) was utilized to assess achievement in mathematics and reading. Descriptions of each follow.

Classroom Assessment Scoring System (CLASS)

Overview of CLASS

“What constitutes quality in [early childhood] classrooms and how it is measured remain in the forefront of discussions about assessing and improving the value of classroom settings for children’s development” (La Paro et al., 2004, p. 420). In fact, the No Child Left Behind Act, a piece of major federal legislation addressing school quality and outcomes, “highlights the importance of quality in early education classrooms and emphasizes accountability and the use of teaching methods that improve student outcomes.” However, until recently, there were no standardized measures available that could accurately or appropriately assess prekindergarten or early elementary classrooms’ contributions to children’s achievement in a system of accountability. A relatively new instrument, The Classroom Assessment Scoring System, developed by Robert Pianta and his colleagues at the University of Virginia, Charlottesville, does address this problem. The CLASS provides a “framework for observing key dimensions of classroom practices, such as emotional and instructional support, that contribute to the quality of the classroom setting from preschool through third grade” (p. 409).

The CLASS has been introduced and used extensively at a variety of levels to measure global classroom quality based on key aspects of quality, dimensions of classroom practices, and teacher-controlled variables, such as emotional and instructional support. The instrument provides a mechanism for examining how teachers interact with
students rather than measuring resources and limitations imposed by variables they cannot control.

Two widely used versions of this instrument were available at the time of the study; one for the preschool level and another for kindergarten through third grade. The K-3 version was utilized in this study. According to Hamre et al. (2006), the CLASS was developed based on research that suggests “interactions between students and adults are the primary mechanism of student development and learning” (p.1). As such, classroom assessment is not based on the materials present, curriculum adoption, safety, or the physical environment in the setting. Rather, the CLASS allows the researcher to focus on “what teachers do with the materials they have and the interactions they have with the students” (p.1).

Howes et al. (2008) also suggested that the “quality of children’s classroom experiences rather than structural features [predict] more growth in children’s academic skills and behaviors.” In other words, they seem to agree that what children do and with whom they engage at school may be more important than the physical environments in which they attend. “This finding, if sustained in future work, has important implications for policy recommendations for quality improvement” at the preschool and elementary levels (p. 47). Perhaps rather than focusing on structural aspects of schools, decision makers will have adequate evidence to shift the weight of accountability emphasis to process variables, which have greater potential to impact positive socio-emotional and academic change and outcomes for children.

The following segment includes a description of the process used for development of the Classroom Assessment Scoring System and provides a comparison to other quality
measures used in early childhood settings. A brief summary of studies used to establish validity and reliability of the instrument is also included.

Development of CLASS

The CLASS was developed to provide a means of accountability for classroom quality based on observable process variables rather than accountability based entirely on assessment of children or structural program features. While the majority of existing assessment systems relies, at least in part, on structural features of the classroom or indicators such as credentialing or class size, the CLASS assesses process variables of quality. These include “the kind of instruction and interactions with adults that occur in prekindergarten and elementary settings, [which] have reliable and detectable effects on children’s achievement and social competence” (La Paro et al., 2004, p. 411). The CLASS dimensions, based on developmental theory and research, examine observed interactions between students and adults rather than structural markers such as curriculum or availability of resources, which are typically adequate and well-organized in public school settings (Hamre et al., 2006).

According to the developers of the CLASS, classroom interactions and practices are variables that can be measured and improved systematically at the classroom level, empowering teachers to increase global classroom quality in three critical domains: Emotional Support, Classroom Organization and Instructional Support. Although structural characteristics, such as teacher credentials, teacher/student ratios, curriculum, physical environment and a number of other variables “provide some sense of resources available for children’s learning, research is inconsistent in uncovering relations between these structural features, classroom quality, and consequently, improved student
outcomes” (LoCasale-Crouch et al., 2007, p. 5). CLASS developers insist that how teachers interact with children and what they do with the resources at their disposal are the real determining factors in program quality (La Paro et al., 2004).

CLASS developers also suggest that the CLASS creates a “common metric and vocabulary” which can be used to describe multiple quality variables across grade levels. These are based on extensive research in the form of literature review. The structure of the instrument, also based on research, divides interactions between students and teachers into three categories or domains: Emotional Support, Classroom Organization and Instructional Support. Within these domains, ten dimensions (or specific indicators) are assessed. The figure below, taken from the CLASS Manual, provides a graphic representation of the domains and dimensions measured by the CLASS (Pianta et al., 2008, p. 2).

Figure 3.1 Overview of Class Domains and Dimensions
Reliability and Validity of the CLASS

Face and construct validity. The technical manual provided with the CLASS claims face and construct validity. Construct validity is based on the method used for development of the instrument. Hamre et al. state,

The CLASS was developed based on an extensive literature review on classroom practices shown to relate to children’s social and academic development in schools. The dimensions were derived from a review of constructs assessed in classroom observation instruments used in child care and elementary school research, literature on effective teaching practices, focus groups, and extensive piloting. Throughout this process, numerous experts in classroom quality and teaching effectiveness have agreed that the CLASS measures aspects of the classroom that are of importance in determining student performance, suggesting considerable face validity. (p. 16)

To provide further evidence of its construct validity, CLASS developers cite its use in multiple large scale studies, which they claim has established validity of the organizational structure for classroom interactions in over 3,000 classrooms (Hamre et al., 2006). In a study designed to field test and establish technical reliability of the CLASS, La Paro et al. (2004) described a number of alternate instruments that provide a means of rating classroom environments on a spectrum of clearly defined elements alleged to index quality at a global level. La Paro and her colleagues compared the CLASS to two of these instruments, the ECERS (Early Childhood Environmental Rating Scale created by Harms, Clifford and Cryer, 1998) and the Snapshot (created by Ritchie,
Howes, Kraft-Sayre, and Weiser, 2002), which measure classroom process quality to provide information about convergent validity of the CLASS.

[They] examined ratings from the CLASS in relation to ratings from two other observational measures used in the study, the ECERS and the Snapshot. Both emotional and instructional support factor scores from the CLASS were related to the ECERS total score, .52, $p < .0001$, and .40, $p < .0001$, respectively. The emotional climate and instructional support factors from the CLASS were most strongly related to the interaction and language reasoning subscales from the ECERS.

Research presented by La Paro et al. (2004) demonstrates correlations between the CLASS and the ECERS, as well as the CLASS and the Snapshot, showing strong correlations between the CLASS scales and ECERS total and subscale scores, particularly on subscales related to process quality. La Paro et al. claim that “CLASS factors are related to ratings from the ECERS in the expected directions and magnitudes” (p. 422), showing stronger correlations to instructional and interactional factors than structural factors. They also identify strong correlations between “the emotional support factor from the CLASS and teacher-child engagement from the Snapshot, but weaker correlations between CLASS emotional support and Snapshot encourages scale” (p. 423). They argue this is likely due to differences in the way the constructs are defined for coding rather than actual differences in factors.

The CLASS manual provides additional evidence that the CLASS is “associated empirically with other measures of similar constructs” (Hamre, et al., 2006, p. 16). Data presented in the Technical Manual of the CLASS, show the results of “analyses
examining relationships between the CLASS (preschool version) and various other measures of classrooms and teachers” (p. 16), again showing positive correlations between the CLASS and the ECERS and Snapshot, with the strongest correlations related to the ECERS. The table provided in the manual also shows negative correlations between the CLASS and the Teacher Depression and Adult-Centered Attitudes scales, showing that teachers who were depressed or held adult-centered attitudes were more likely to have lower quality ratings on the CLASS.

**Predictive validity.** The CLASS was developed to measure “classroom-level processes that are directly associated with children’s performance” (Hamre, et al., 2006). In a study conducted by Howes, Burchinal, Pianta, Bryant, Early, Clifford, and Barbarin (2008), researchers established a connection between academic outcomes and quality of instruction and teacher-child relationships. Using data from two studies, the National Center for Early Development and Learning (NCDEL) Multi-State Study of Pre-Kindergarten and the State-Wide Early Education Programs Study (SWEEP), researchers conducted hierarchical linear model (HLM) analysis to assess whether children achieved higher academic gains when exposed to higher quality programs. Results from this study showed “modest associations between some dimensions of classroom quality and gains in academic skills” as measured by the Woodcock-Johnson Achievement test (p. 39). However, the researchers cite prior work by Hamre and Pianta (2005) in which higher scores on the CLASS were related to “greater gains on achievement tests” (Howes et al., 2008). Findings from this study suggest that at-risk students in first grade classrooms with high levels of emotional and instructional support did as well as their peers who were not labeled at risk. These two studies support the predictive validity of the CLASS.
**Scorer reliability.** According the developers of the CLASS, “reliability refers to the degree to which an instrument is free from random error associated with the process of measuring the construct of interest” (Pianta et al., 2008, p. 99). To increase reliability of scoring for this study, the researcher participated in two days of training using materials that provided “a clear and comprehensive understanding of the instrument’s purposes and procedures” (p. 99). These training sessions were conducted by “gold standard” scorers at CASTL (Center for Advanced Study of Teaching and Learning) at the University of Virginia, Charlottesville, who were trained and certified by the instrument’s developers. Master codes were established by consensus coding by developers until agreement was reached for training and testing videos.

Reliability training also included watching multiple videotaped segments of classroom interactions that were consensus coded by at least three master CLASS coders. These consensus ratings established a standard by which to judge future observations. After training, the researcher watched additional videotaped classroom segments which she coded and testers compared to master codes to determine inter-rater reliability compared to master coders. The researcher successfully demonstrated 80% reliability as a scorer as compared to master codes. This means she scored video segments of classroom interactions within one point of master coders at least 80% of the time.

According to developers, reliability of scores is also supported by providing multiple observation cycles. Prior research demonstrated that four cycles (for a total of approximately two hours) in each setting provided a strong representative sample of classrooms, providing a high degree of internal consistency for CLASS scores across the cycles (Pianta et al., 2008).
Analysis of scoring cycles was conducted in a variety of ways to demonstrate stability of scoring across time. Researchers examined the degree to which the scores for the first four cycles of observation correlated with the final scores in a study with a preschool (Multi-State Prekindergarten Study) and a third grade (4R’s Study) sample. Correlations ranged from .84 to .91, demonstrating high levels of reliability across cycles. In addition, the researchers assessed the stability of scores by measuring the “degree of consistency of CLASS scores across two, three, and four cycles” with the same samples. Results showed correlations from .63 to .91 across all dimensions (Hamre et al., 2006, p. 11).

The researcher followed the recommended model for CLASS observations and assessment, completing four cycles of observation in all but one kindergarten classroom. In this exception, the classroom teacher went on leave after two observation cycles. Rather than eliminating this teacher and his/her students from the study, the researcher utilized the data collected and included the teacher in the study. Scores obtained by the researcher in the current study were consistent across observation cycles and days of the week.

In sum, the CLASS was selected for this study over other measures of classroom quality, particularly the ECERS, in spite of its wide use in evaluation of early childhood programs because the CLASS has high validity and reliability and because of its focus on process variables rather than structural variables.
Overview of the NWEA MAP

In order to assess student academic achievement, the researcher utilized scores from NWEA MAP Achievement Tests, computer adaptive version. NWEA MAP was available for two levels; the MAP Primary version, designed for use at the kindergarten through second grade levels, and the MAP 2 – 10 version, which was designed for use with children at grades two through ten. Kindergarten students were assessed using the Primary Version. Although the MAP 2-10 was available for students who reached the maximum score on the Primary version, none of the students were assessed utilizing the 2-10 version.

A general description of the tests, taken from Northwest Evaluation Association’s website (www.nwea.org) follows:

Measures of Academic Progress are electronically administered and scored achievement tests designed to measure growth in student learning for individual students, classrooms, schools, and districts. The tests provide accurate and immediate scores to help teachers plan instructional programs, place new students in the appropriate courses, and screen students for special programs. MAP is a computerized adaptive testing system that tailors tests to a student’s achievement level. Each student takes a test that is dynamically developed for him or her as the test is being administered. The program instantly analyzes the student’s response to each test question and, based on how well the student has answered all previous questions, selects a question of appropriate difficulty to display next. (Northwest Evaluation Association: FAQ, 2008)
The assessments available in the MAP for Primary Grades “include Early Literacy and Early Numeracy Screening (diagnostic) tests [and] Skills Checklist (diagnostic) tests” in addition to the above mentioned Survey with Goals (adaptive) tests in Reading and Mathematics. Scores for diagnostic tests are reported as percentages and number correct, and survey with goals test scores are reported in Rasch Unit (RIT) scores along with goal score levels to help determine instructional levels. NWEA MAP for Primary Grades is designed specifically for early learners (Northwest Evaluation Association: Overview, 2008). The Survey with Goals tests were utilized for this study. NWEA’s website states:

MAP for Primary Grades meet the unique needs of early learners by utilizing advanced technology to display interactive visuals and audio for beginning readers. For example, the computer automatically plays audio instructions to the student, eliminating the challenges of early learners who cannot read. Students only need to be able to click a mouse to perform an action; there is no need to hold the mouse button down while moving it (Northwest Evaluation Association: Features, 2008).

In addition to the above features, the MAP for Primary Grades offers a Spanish version for students not proficient in English (Northwest Evaluation Association: Features, 2008). The district utilizes the Spanish version of the mathematics test for children whose first language is Spanish and who have not yet developed English proficiency. However, there were no instances of the use of the Spanish language version
during this study as there were no children identified with limited English proficiency in the study group.

NWEA MAP are “state-aligned computerized adaptive tests that accurately reflect the instructional level of each student and measure growth over time” (NWEA).

According to the official NWEA web page, “MAP tests provide highly accurate results that can be used to:

- Identify the skills and concepts individual students have learned.
- Diagnose instructional needs.
- Monitor academic growth over time.
- Make data-driven decisions at the classroom, school and district levels.
- Place new students into appropriate instructional programs.” (www.nwea.org)

Development of the NWEA MAP

NWEA MAP tests are created from a test bank of more than 15,000 items, with hundreds of new items added each year. According to the Northwest Evaluation Association, these items are developed by teachers who have been thoroughly trained in NWEA’s item-writing process.

Each potential item must then pass a rigorous bias and content review, which is followed by field-testing, and the subsequent strict statistical screening procedures are calibrated for difficulty and assigned the appropriate value on the RIT scale.

These items become part of the continually expanding item bank. (NWEA: Research-Based Accuracy, n.d.)
Reliability of NWEA MAP

The extensive item bank of questions used on the NWEA Measures of Academic Progress (MAP) tests have been developed over a substantial period of time. This has given staff charged with statistical analysis abundant opportunity to establish the reliability of the tests. The result has been the collection of a significant amount of reliability evidence over time. (NWEA: Reliability, n.d.)

Statistical data to support claims of reliability and validity of MAP tests are found on the NWEA legacy support website, which houses research related to the tests. According to this report, ―NWEA’s approach to test-retest reliability poses a more rigorous test of reliability‖ than is typical because of the extended time between testing and retesting (NWEA, 2004, para. 5). While the standard process is to test and retest using the same test within a few weeks of each other, NWEA utilizes a “mix between test-retest reliability and a type of parallel forms of reliability, both of which are spread across seven to twelve months…” (NWEA, 2004, para. 5). In spite of this more rigorous methodology, most reliability coefficients are in the mid-.80’s to the low .90’s suggesting strong reliability of NWEA MAP.

Validity of NWEA MAP

To insure content validity, the developers of the Measures of Academic Progress map existing content standards from a district or state into a test blueprint, and then match test items to the content standards and level of the test being created. Then the tests are compared to other established achievement tests to determine if they accurately measure the identified constructs. (NWEA, 2004)
Validity of the NWEA MAP has been assessed in the form of concurrent validity. This shows how well the NWEA measures achievement in terms of RIT scale scores in subject areas. When comparing NWEA outcomes to several other standardized achievement tests (Arizona Instrument to Measure Standards; Colorado Student Assessment Program; Illinois Standards Achievement Tests; Indiana Statewide Testing for Educational Progress-Plus; Iowa Tests of Basic Skills; Minnesota Comprehensive Assessment and Basic Skills Test; Nevada Criterion Referenced Assessment; Palmetto Achievement Challenge Tests; Stanford Achievement Test, 9th Edition; Texas Assessment of Knowledge and Skills; Washington Assessment of Student Learning and ALT; and Wyoming Comprehensive Assessment System and ALT), reliability coefficients are between .69 and .89 across tests and subject areas. However, all of the comparison tests begin at the second grade level. No data is provided for the primary level MAP which was used in this study.

Procedure

Data Collection

Measuring Academic Achievement

NWEA Pre-test. Data collection for this study took place in multiple stages. Initial data collection occurred between August 25 and September 19, 2008 (fall test), following student entry into the kindergarten classroom. During the first few weeks of school, the district assessed all kindergarten students using the NWEA MAP Primary Version as part of their normal assessment process for mathematics and reading. Achievement scores taken from the mathematics and reading subtests were archived, coded, and provided to
the researcher after study approval. These scores were used to establish baseline data to demonstrate academic growth over time.

*NWEA Retests*. During the final stages of data collection, all students were reassessed using the NWEA MAP Primary version mathematics and reading tests. This took place between December 1, 2008 and January 9, 2009 (winter test). An expanded time frame was required to accommodate the winter holiday. The final retest took place between March 30 and May 1, 2009 (spring test). Again, an expanded window was established to compensate for the district’s spring break. These retest scores were recorded and archived by the district, then coded to student numbers matching initial assessment. Coded scores were made available to the researcher in the form of PowerSchool® spread sheets for analysis after study approval and completion of CLASS observations (District Director of Special Programs, personal communication, March 18, 2010.).

Initially, the researcher intended to utilize only fall and winter scores to determine academic growth in mathematics and reading. However, concern over validity of fall test scores for children due to their lack of school experience and developmental readiness for the standardized computer testing format, the researcher decided to include spring scores to provide additional strength to the study.

*District Demographics*

In addition to achievement data, demographic data for student and teacher participants were collected and coded by the district during the fall. Demographic information collected for students included: identification number; gender; ethnicity; free, reduced or full-pay meal status; Title I designation; special needs designation (high
ability or other exceptionality); limited English proficiency; school and teacher of record. Teacher demographics collected included: gender, level of education, and number of years teaching. Classroom demographics included class size and program format (all-day every day, alternating day, or K+).

Classroom Visits

The next level of data collection consisted of four randomly scheduled classroom observation cycles conducted by the researcher in each of the twenty-eight district kindergarten classes. Randomization took place by random cluster sampling at the school level. The researcher randomly selected a school for each observation day and visited all kindergarten classrooms located at the selected school on that date, as time allowed. This process continued until all classes had been observed for at least four cycles. Teachers were not informed in advance of visits that took place between January and April of 2009, using the CLASS to record observations.

Each observation began with the researcher entering the classroom and determining where to position herself for minimum distraction. Typically, a table near the rear of the classroom provided the researcher ample opportunity to hear and see what was occurring without interfering with the activity taking place. The researcher then began making observations and recording behaviors related to each of the dimensions on the classroom assessment record sheet. Scoring for the CLASS took place immediately following each observation cycle utilizing an observation score sheet that incorporated a seven-point Likert-type scale. Immediate scoring allowed the observer to assign a score for each of the ten dimensions representing quality as observed during each observation
cycle (Pianta et al., 2008). Individual score sheets for each cycle were retained for later analysis.

**Analysis of Data**

*NWEA Scores*

RIT (Rasch Unit) scores on the NWEA MAP Primary Version indicate normed achievement levels for mathematics and reading. These archived scores from the initial administration and retests were compared to determine academic growth for students during the study period. Fall to winter growth, winter to spring growth, and overall growth for the full academic year were recorded and analyzed. Adjustments were made for covariates at the school, class and individual levels to account for outcomes beyond those measured by the CLASS. These covariates are discussed in the section describing comparison of quality and achievement.

*CLASS Quality Ratings*

After all observation and scoring cycles using the CLASS were completed, composite scores for each dimension were calculated. To attain composite scores, individual dimension scores were averaged across all cycles. For all but one class, this included scores from four cycles of observation. In one case, only two cycles were completed because the classroom teacher took a leave of absence. Because the teacher left very late in the spring semester, the spring NWEA retest had already been administered, so achievement scores were not affected. In this case, the two completed cycles were averaged to create a composite score for this class. Once all average dimension scores were calculated, composite domain scores were determined for Emotional Support, Classroom Organization, and Instructional Support. These were
calculated using a summary scoring sheet and averaging scores within each domain. This formula for calculation of domain composite scores requires recording and averaging scores as positive numbers, so scores for negative climate were reverse scored for the calculation. The higher a score on any domain or dimension the more positively it is viewed. This followed the recommendations included in the CLASS manual (Pianta et al., 2008).

Assessing the relationship between Teaching Quality and Student Achievement

After all data were collected, the researcher utilized SPSS Graduate Pack 15.0 for Windows (2006), a statistical analysis program, to analyze data. Prior to the primary analysis of data, the researcher examined the normality of the distribution of NWEA scores using QQ Plots. Normality is an assumption of the statistical model used to address the researcher’s questions. Fall to winter, winter to spring, and overall growth scores were checked for both reading and mathematics. Normal distributions were displayed, so the researcher proceeded with analysis of the data described next.

A hierarchal linear model [HLM] for analysis was used to answer the research questions and to account for the multilevel structure of the data collected within school settings. The researcher compared classroom quality scores to student achievement in mathematics and reading. Because a great many variables may potentially impact student achievement, the HLM allowed the researcher to look at many variables and determine which had an impact on achievement, in addition to classroom quality. This multilevel model also allowed for the inclusion of covariates at the school, class, teacher, and student levels. Each dependent variable (change in reading scores fall to winter, change in reading scores winter to spring, change in reading scores fall to spring, change in math
scores fall to winter, change in math scores winter to spring, and change in math scores fall to spring) was assessed after building the model isolating main effects of degree, gender, lunch status, kindergarten program and years of teacher experience. The random effect of teacher nested within school was also included in the model.

Parameter estimates, covariance parameters and estimated marginal means were calculated for degree, gender, lunch status, and kindergarten program. These provided additional information about the range of scores, and differences among different groups of students (by gender, program, class, etc.). This analysis was repeated for each of the ten measures (dimensions) of classroom quality as defined by the CLASS and was repeated for each growth period of the NWEA MAP (fall to winter, winter to spring, and overall growth) for mathematics and reading. In all, this required 70 analyses.

CLASS variables were all run separately rather than as a group due to the colinear nature of the scores. A full model including all dimensions would not run properly. In spite of the elevated Type I Error rating created when multiple analyses are run, the researcher chose to stick with an alpha level of .05 for all statistical comparisons. The exploratory nature of this research makes this reasonable; however, it is probable that some results may appear significant within the study sample, but not the whole population. The researcher accepts this limitation because her primary interest relates to the local school district and further research is indicated to address inconsistencies with prior research before generalizations might apply.

Keller (2006) stated that the HLM model is appropriate “when situations exist whereby the subjects of an analysis are somehow contained within another group” (p. 137). In this study, teacher is nested within the school by degree and students (scores) are
nested within the class (teacher within school). Ma, Ma, and Bradley (2008) provided an explanation. They stated, “School education systems provide an obvious example of hierarchal structure, with students nested within classes, classes nested within schools, schools nested within districts, and so on” (p. 63). For example, this nested model allowed the researcher to identify and assess the impact of individual student variables, teacher- and class-level variables, and school-level variables on student achievement and allowed for model building to focus on those factors which were determined to be significant in addition to the primary variable of interest, classroom quality.

Belsky et al. (2007) reminded us that “such correlational research does not allow strong inferences regarding causation, as efforts to control confounding factors can never insure that all important ‘third variables’ or alternative explanations have been taken into account” (p. 681). For example, it was not possible to determine the effect concurrent home and after school experiences had on “developmental trajectories” (p. 683). Smith (1996) acknowledged these additional variables, which may have affected children’s development and warned that one must consider such external factors when assessing program quality. The researcher recognizes the inability to be certain all covariates have been considered. However, utilization of hierarchal linear regression allowed her to identify and control for several student level extraneous variables such as student gender, socio-economic status, and program format and teacher-level variables such as gender, level of education and years teaching.
CHAPTER 4: RESULTS

Overview

This study was designed to address a current concern in the local school district, how to improve student achievement. Like most, the local district is dealing with pressure from the state to meet NCLB mandates “to achieve unprecedented educational progress…ensuring that all students—and all subgroups—meet the state’s proficiency goals” (Sunderman & Orfield, 2007, p. 138). To address these issues, the researcher used a hierarchal linear model (HLM) to analyze the relationship between classroom quality and student achievement in district kindergarten classrooms while controlling for other variables (teacher degree, student gender, student lunch status, kindergarten program variance, teachers’ years of experience, and school attended). This allowed the researcher to identify and assess classroom quality that might lead to higher achievement for all students. Classroom quality was assessed using a standardized assessment instrument (CLASS). Student academic achievement in mathematics and reading, as measured by NWEA MAP, was compared with classroom quality across three broad domains (Emotional Support, Classroom Organization, and Instructional Support). Within these domains, ten specific quality dimensions (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Student Perspectives, Behavior Management, Productivity, Instructional Learning Formats, Concept Development, Quality of Feedback, and Language Modeling) were assessed and analyzed to determine if classroom quality and student achievement in mathematics and reading were related.
Research Questions

This study addressed four questions about the relationship between kindergarten achievement and classroom quality:

Research Question 1: Is kindergarten achievement growth in mathematics related to global classroom quality as measured by the CLASS?

Research Question 2: Is kindergarten achievement growth in reading related to global classroom quality as measured by the CLASS?

Research Question 3: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in mathematics?

Research Question 4: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in reading?

Descriptive Statistics for CLASS Domains and Dimensions

Prior to reporting the results of the HLM analysis, descriptive statistics for the variables used in this study are presented. The Classroom Assessment Scoring System (CLASS) is an instrument designed to assess classroom process quality in three domains: Emotional Support, Classroom Organization, and Instructional Support. Each of these domains is further divided into more specific dimensions of classroom quality. Descriptive statistics for both broad domains, specific dimensions, and a global score of classroom quality are presented in Table 4.1, which follows:
Table 4.1
Descriptive Statistics for CLASS

<table>
<thead>
<tr>
<th>Domains, Dimensions, and Global Quality</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite Emotional Support</td>
<td>3.06</td>
<td>7.00</td>
<td>5.5052</td>
<td>1.18150</td>
</tr>
<tr>
<td>Positive Climate</td>
<td>2.50</td>
<td>7.00</td>
<td>5.7859</td>
<td>1.24307</td>
</tr>
<tr>
<td>Negative Climate</td>
<td>3.50</td>
<td>7.00</td>
<td>6.2649</td>
<td>1.00900</td>
</tr>
<tr>
<td>Teacher Sensitivity</td>
<td>2.75</td>
<td>7.00</td>
<td>5.1869</td>
<td>1.35977</td>
</tr>
<tr>
<td>Regard for Student Perspectives</td>
<td>2.75</td>
<td>7.00</td>
<td>4.7803</td>
<td>1.40536</td>
</tr>
<tr>
<td>Composite Classroom Organization</td>
<td>2.92</td>
<td>6.92</td>
<td>5.5353</td>
<td>1.00666</td>
</tr>
<tr>
<td>Behavior Management</td>
<td>3.00</td>
<td>7.00</td>
<td>5.8806</td>
<td>1.03422</td>
</tr>
<tr>
<td>Productivity</td>
<td>3.00</td>
<td>7.00</td>
<td>5.6485</td>
<td>1.02960</td>
</tr>
<tr>
<td>Instructional Learning Format</td>
<td>2.75</td>
<td>7.00</td>
<td>5.0767</td>
<td>1.14340</td>
</tr>
<tr>
<td>Composite Instructional Support</td>
<td>2.00</td>
<td>6.75</td>
<td>4.1503</td>
<td>1.54809</td>
</tr>
<tr>
<td>Concept Development</td>
<td>2.00</td>
<td>6.75</td>
<td>3.9499</td>
<td>1.67055</td>
</tr>
<tr>
<td>Quality of Feedback</td>
<td>1.75</td>
<td>7.00</td>
<td>4.2506</td>
<td>1.57460</td>
</tr>
<tr>
<td>Language Modeling</td>
<td>1.50</td>
<td>7.00</td>
<td>4.2512</td>
<td>1.49264</td>
</tr>
<tr>
<td>Global Quality Score</td>
<td>2.75</td>
<td>6.80</td>
<td>5.0889</td>
<td>1.21566</td>
</tr>
</tbody>
</table>

Note: Descriptive statistics calculated using SPSS Graduate Pack 15.0 for Windows. Range of possible scores on the CLASS is 1.0 to 7.0 on a Likert-type scale. n = 28 for all dimensions.
The table above presents descriptive statistics for the CLASS. Possible scores on each dimension range from 1 to 7 on a Likert-type scale. Scores of 1 and 2 are considered low scores; 3, 4 or 5 are considered middle-range scores; and a score of 6 or 7 is considered a high score for each dimension (Pianta, La Paro, & Hamre, 2008). Mean scores may include scores that fall between the whole numbers. However, the testing manual does not identify a clear cut-off point for low, medium and high scores between the wholes. Higher scores are considered better for all dimensions except Negative Climate. To allow for calculations of mean composite (domain) scores, the scores for Negative Climate are reverse scored so that the higher a score on any domain or dimension the more positively it is viewed. For all scores identified below, mean scores with (standard deviations) are provided.

To calculate a global quality score for each class, the researcher calculated a mean score for each dimension of quality assessed using the CLASS. The means for these ten dimension scores were then added together, and divided by ten to achieve an overall (or global) mean score for classroom quality. Although there is no prior research indicating calculation in this manner, the researcher extended the process used to calculate domain composite scores.

Scores on the CLASS range from 1.50 for Language Modeling, to 7.0 on several dimensions (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Student Perspectives, Behavior Management, Productivity, Instructional Learning Format, Quality of Feedback, and Language Modeling). Mean scores for all dimensions were within the medium range, with the exception of Negative Climate, which had a mean score of 6.26 (1.01). The highest mean scores were attained within the Emotional Support and
Classroom Organization domains. All dimensions within these domains had mean scores above 5.5, with standard deviations below 1.4 with the exception of Regard for Student Perspectives, which had a mean score of 4.78 (1.4). This may indicate that kindergarten classes in the study district tend to provide high levels of emotional support and classroom organization, but that teacher-centered practices often prevail.

While Classroom Organization and Emotional Support scored in the high medium range, the lowest scores attained on the CLASS in this district were within the Instructional Support Domain. The mean composite score was 4.15 (1.5), which is in the lower medium range; all dimensions scored 4.25 or less with standard deviation of 1.67 or less. This may suggest that district kindergarten classrooms, while positive and organized, may provide less instructional support than would be ideal.

Descriptive Statistics for NWEA MAP
Scores in Mathematics and Reading

Overview

Mathematics and reading achievement for all students in the school corporation were assessed three times during the academic school year using the the NWEA MAP Primary Version (computer adaptive version). The first administration of the test was conducted during a “testing window” from August 25, 2008 through September 19, 2008, which was within the first few weeks of of school. This was done to establish baseline (pretest) scores in each of the subject areas tested. Follow-up testing (posttests) in the winter (December 1, 2008 through January 9, 2009) and spring (March 30, 2009 through May 1, 2009) were then conducted to measure achievement gains in mathematics and reading. (District Director of Special Programs, personal communication, March 18,
Achievement gains in mathematics were calculated by subtracting initial (pretest) mathematics RIT scores from follow-up (posttest) scores to establish achievement gain during the study period. Fall to winter gain, winter to spring gain, and overall gain from fall to spring were measured.

In addition to data from the current study, norming data for the NWEA MAP are provided for comparison purposes. According to the 2008 Normative Data Report, the 2008 NWEA RIT Scale Norms Study examined 54,000 MAP results from children in kindergarten and first grade who were tested between the fall of 2006 through spring of 2007. All scores were included in the norming group because the sample was not large enough to support a stratified sample (NWEA: Normative, 2008).

Table 4.2 summarizes standard scores on the NWEA in mathematics and reading for the 2008-2009 school year. It provides mean standard scores for all district kindergarten students in reading and mathematics for the fall, winter, and spring administrations of the MAP. This table also includes mean standard scores of the norming population (2008) for comparison purposes. These scores are reported as RIT (Rasch Unit) scores. Scores for students in the local district were comparable (within a few points) to those of the norm group in all areas.
Table 4.2

Descriptive Statistics for NWEA MAP Standard Scores in Mathematics and Reading

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean (SD)</th>
<th>Norm Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>403</td>
<td>116</td>
<td>173</td>
<td>143.83 (10.597)</td>
<td>149.5</td>
</tr>
<tr>
<td>Winter</td>
<td>403</td>
<td>119</td>
<td>179</td>
<td>152.58 (10.707)</td>
<td>153.1</td>
</tr>
<tr>
<td>Spring</td>
<td>404</td>
<td>118</td>
<td>191</td>
<td>161.89 (11.655)</td>
<td>158.1</td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>403</td>
<td>110</td>
<td>170</td>
<td>142.09 (8.806)</td>
<td>147.6</td>
</tr>
<tr>
<td>Winter</td>
<td>404</td>
<td>128</td>
<td>183</td>
<td>152.58 (9.184)</td>
<td>152.4</td>
</tr>
<tr>
<td>Spring</td>
<td>403</td>
<td>134</td>
<td>193</td>
<td>159.52 (9.927)</td>
<td>156.3</td>
</tr>
</tbody>
</table>

Valid N (listwise) 403

Note. All data for the current study was compiled utilizing SPSS Graduate Pack 15.0. Scores on the NWEA MAP are recorded as RIT (Rasch Unit Scores). Standard Norm scores were attained from NWEA: Normative Data 2008; no standard deviations were provided. Min = Minimum; Max = Maximum; SD = Standard Deviation.
The data provided in Table 4.3 provide additional descriptive statistics for the NWEA MAP. Data include minimum and maximum gain scores for each growth period in mathematics and reading for the study group and the norming population: fall to winter, winter to spring, and over all growth from fall to spring. Negative minimum scores indicate that some students achieved a lower score on the posttest than they did on the pretest.

Local district mean growth scores were higher in mathematics and reading than those of the norm group across all growth periods. However, when standard deviations are considered for the study group’s scores, they appear to be comparable. This information is also shared to provide a subjective basis for comparison of the scores in the current study district.
Table 4.3

Descriptive Statistics for NWEA MAP Gain Scores in Mathematics and Reading

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Minimum Gain</th>
<th>Maximum Gain</th>
<th>Mean Gain</th>
<th>Standard Deviation</th>
<th>Norm Mean Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall to winter</td>
<td>403</td>
<td>-15.00</td>
<td>30.00</td>
<td>8.7469</td>
<td>7.03937</td>
<td>3.6</td>
</tr>
<tr>
<td>Winter to spring</td>
<td>403</td>
<td>-10.00</td>
<td>39.00</td>
<td>9.3499</td>
<td>7.06186</td>
<td>5.0</td>
</tr>
<tr>
<td>Fall to spring</td>
<td>404</td>
<td>-13.00</td>
<td>45.00</td>
<td>18.0619</td>
<td>8.17177</td>
<td>8.6</td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall to winter</td>
<td>403</td>
<td>-8.00</td>
<td>36.00</td>
<td>10.6650</td>
<td>6.45955</td>
<td>5.7</td>
</tr>
<tr>
<td>Winter to spring</td>
<td>404</td>
<td>-14.00</td>
<td>25.00</td>
<td>.7916</td>
<td>6.19229</td>
<td>3.9</td>
</tr>
<tr>
<td>Fall to spring</td>
<td>403</td>
<td>-8.00</td>
<td>42.00</td>
<td>17.4257</td>
<td>7.40243</td>
<td>9.6</td>
</tr>
<tr>
<td><strong>Valid N (listwise)</strong></td>
<td>403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. All data was compiled utilizing SPSS Graduate Pack 15.0. Scores on the NWEA MAP are recorded as RIT (Rasch Unit Scores). Negative scores indicate that some students achieved lower scores on retests than on the pretests.
Summary

For the 2008-2009 school year, local achievement in mathematics and reading was comparable to that of the norming group. Local mean scores were within a few points of the norm group’s mean scores in all areas when standard deviations were considered. Sixty eight percent of district kindergarteners scored at or higher than the median score when compared to the NWEA Norm Study in Reading and sixty four percent of district kindergarteners scored at or higher than the median score RIT score in mathematics (District Director of Special Programs, personal communication, March 18, 2010). However, there was a wide range of variability in scores with some students scoring much higher than the mean and others scoring lower on posttests than on the pretest.

Significance of Classroom Quality on Reading and Mathematics Achievement

The researcher posed four research questions, each of which appear below, followed by a report on the analyses used to address them. Data from these analyses appear in Table 4.4, which follows the research questions.

Research Question 1: Is kindergarten achievement growth in mathematics related to global classroom quality as measured by the CLASS?

In order to determine whether kindergarten achievement in mathematics was related to classroom quality, the researcher first examined the normality of the distribution of mathematics scores for students by means of QQ Plot. Normality is the key assumption underlying the statistical model used here. The QQ Plots were evaluated for each of the growth periods: fall to winter, winter to spring, and overall growth from
fall to spring. In all three cases, scores were normally distributed, so the researcher continued to examine the relationship of student achievement (NWEA MAP growth scores) to global classroom quality (composite CLASS scores).

A hierarchal linear model (HLM) was used to adjust for the following fixed effects: teacher degree, student gender, student lunch status, kindergarten program, and teacher’s years of experience. The researcher also controlled for the effects of the student’s school, teacher and teacher’s degree [school + teacher (school * degree)].

As shown in Table 4.4, results of the HLM analyses allowed the researcher to reject the null hypothesis for Question 1. Global classroom quality showed a significant positive slope of .835 at the .05 alpha level ($\alpha$), when compared to mathematics achievement growth from fall to winter. However, growth scores for winter to spring and overall mathematics achievement (fall to spring) were not significantly related to global classroom quality at $\alpha$.05.

To determine if other covariables (teacher degree, student gender, student lunch pay status, kindergarten program, and teacher years of experience) may have affected outcomes for this analysis, the researcher checked Estimates of Fixed Effect for to determine if any other variables were significant at the 0.05 level. In this case, none of the covariables were significantly related.

**Research Question 2: Is kindergarten achievement growth in reading related to global classroom quality as measured by the CLASS?**

In order to determine whether kindergarten achievement in reading was related to global classroom quality as measured by the CLASS, the researcher examined the normality of the distribution of reading scores for students by means of QQ Plot, as was
done for mathematics scores. In all three cases (fall to winter, winter to spring and fall to spring), scores were normally distributed, so the researcher continued to examine the relationship of student achievement (NWEA MAP growth scores) in reading to global classroom quality (composite CLASS scores). To do this, the same process was used as for Question 1, substituting reading scores for mathematics.

As shown in Table 4.4, results of the analyses supported the null hypothesis for reading. At the .05 alpha level, reading achievement did not show a significant positive relationship to global classroom quality as measured by the CLASS during the fall to winter growth period, the winter to spring growth period or for overall reading growth.

Research Question 3: Which specific indicators, as measured by the CLASS, are related to kindergarten achievement growth in mathematics?

To answer this question, each domain and dimension of classroom quality was analyzed separately, while adjusting for the following fixed effects: teacher degree, student gender, student lunch status, kindergarten program, and teacher’s years of experience. The researcher also controlled for the effects of the student’s school, teacher and teacher’s degree [school + teacher (school * degree)].

As shown in Table 4.4, results of the analyses allowed the researcher to reject the null hypothesis for Question 3. At the .05 alpha level, mathematics achievement showed a significant positive relationship to Classroom Organization, one of the CLASS domains, as well as two of the subscales within this domain. Both Productivity and Instructional Learning Formats showed strong positive relationships, with higher scores on the CLASS relating to greater growth in mathematics achievement during the fall to winter growth period.
While the composite score for the Emotional Support domain did not show a significant relationship to mathematics achievement during the fall to winter growth period, one of the subscales, Regard for Student Perspectives, did show a positive relationship at a 0.05.

The final CLASS domain, Instructional Support, showed the strongest correlations between classroom quality indicators and mathematics achievement during the fall to winter growth period with an alpha below .001. Within the domain, all three subscales also showed significant correlations between dimensions of classroom quality and mathematics achievement. All three of these were significant at the .01 level. Slope coefficients for these outcomes can be seen in Table 4.4.

While fall to winter mathematics growth scores showed significant correlations to several aspects of classroom quality, winter to spring growth scores and overall growth scores in mathematics were not significantly related to classroom quality within any of the domains or dimensions of the CLASS. Slope coefficients for all of the significant relationships were positive and ranged from .702 to .874. These along with all other slope coefficients may be reviewed in Table 4.4.

As in Question One, the researcher checked the Estimates of Fixed Effects for each indicator of classroom quality (domains and dimension) to determine if any of the other variables had a significant effect on the outcomes. Two of the categorical variables, gender and kindergarten program were significant in a number of cases. For every subscale in which there was a significant relationship between quality and achievement in mathematics, gender was also significant (or very close) at the 0.05 level. In each case, mean male growth scores were at least one point higher than female mean growth scores.
In two cases, Concept Development and Language Modeling, kindergarten program had a significant impact on outcomes at the 0.05 level. In both cases, scores from students attending alternating-day programs had higher mean growth scores (9.760 and 10.132, respectively) than did children attending all-day every-day programs (8.458 and 8.759). Children enrolled in K+ classes had the lowest mean growth scores comparatively (6.444 and 6.568). These scores are reported in Table 4.5.

None of the other fixed effects were significantly related to the outcome for any of the variables of interest (teacher degree, student lunch pay status, or teacher’s years of experience.

Research Question 4: Which specific indicators, as measured by the CLASS, are related to kindergarten achievement growth in reading?

To determine whether classroom quality dimensions have a significant relationship to reading achievement, each domain and dimension of classroom quality was again analyzed separately, while adjusting and controlling for the same effects as for mathematics.

As shown in Table 4.4, results of the analyses supported the null hypothesis in all instances for reading. At the .05 alpha level, reading achievement did not show a significant positive relationship to classroom quality as measured by the CLASS during the fall to winter growth period, the winter to spring growth period or for overall reading growth. Slope coefficients for reading achievement also appear in Table 4.4 which follows.
Table 4.4

**Significance of Classroom Quality – Slope Coefficients**

<table>
<thead>
<tr>
<th>CLASS dimensions</th>
<th>Math fall to winter</th>
<th>Math winter to spring</th>
<th>Math fall to spring</th>
<th>Reading fall to winter</th>
<th>Reading winter to spring</th>
<th>Reading fall to spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite domain–Emotional support</td>
<td>.675</td>
<td>-.198</td>
<td>.290</td>
<td>.262</td>
<td>1.289</td>
<td>.491</td>
</tr>
<tr>
<td>Positive climate</td>
<td>.554</td>
<td>.056</td>
<td>.368</td>
<td>.394</td>
<td>.411</td>
<td>.468</td>
</tr>
<tr>
<td>Negative climate</td>
<td>.534</td>
<td>.242</td>
<td>.282</td>
<td>-.226</td>
<td>.725</td>
<td>.114</td>
</tr>
<tr>
<td>Teacher sensitivity</td>
<td>.605</td>
<td>-.357</td>
<td>.146</td>
<td>.363</td>
<td>.522</td>
<td>.626</td>
</tr>
<tr>
<td>Regard for student perspectives</td>
<td>.702</td>
<td>-.434</td>
<td>.243</td>
<td>.185</td>
<td>.363</td>
<td>.412</td>
</tr>
<tr>
<td>Composite domain–Classroom organization</td>
<td>.874</td>
<td>-.700</td>
<td>.221</td>
<td>.554</td>
<td>.382</td>
<td>.810</td>
</tr>
<tr>
<td>Behavior management</td>
<td>.660</td>
<td>-.365</td>
<td>-.023</td>
<td>.253</td>
<td>.704</td>
<td>.591</td>
</tr>
<tr>
<td>Productivity</td>
<td>.794</td>
<td>-.734</td>
<td>.211</td>
<td>.577</td>
<td>.152</td>
<td>.665</td>
</tr>
<tr>
<td>Instructional learning formats</td>
<td>.871</td>
<td>-.653</td>
<td>.320</td>
<td>.578</td>
<td>.208</td>
<td>.761</td>
</tr>
<tr>
<td>Composite Domain –Instructional support</td>
<td>.833</td>
<td>-.483</td>
<td>.428</td>
<td>.431</td>
<td>.138</td>
<td>.536</td>
</tr>
<tr>
<td>Concept development</td>
<td><strong>.747</strong></td>
<td>-.472</td>
<td>.287</td>
<td>.345</td>
<td>.107</td>
<td>.464</td>
</tr>
<tr>
<td>Quality of feedback</td>
<td><strong>.808</strong></td>
<td>-.348</td>
<td>.561</td>
<td>.457</td>
<td>.097</td>
<td>.539</td>
</tr>
<tr>
<td>Language modeling</td>
<td><strong>.833</strong></td>
<td>-.576</td>
<td>.381</td>
<td>.457</td>
<td>.192</td>
<td>.566</td>
</tr>
<tr>
<td>Global Classroom Quality</td>
<td><strong>.835</strong></td>
<td>-.369</td>
<td>.477</td>
<td>.409</td>
<td>.617</td>
<td>.390</td>
</tr>
</tbody>
</table>

*p < 0.05; **p < 0.01; ***p < 0.001
Calculations derived from SPSS – Fixed effects output for Class domains and dimensions.
Table 4.5

*Estimated Marginal Means for Significant Fixed Effects*

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
<th>Mean Male Growth Score</th>
<th>Mean Female Growth Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emotional Support</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regard for student perspectives</td>
<td>.051</td>
<td>8.911</td>
<td>7.534</td>
</tr>
<tr>
<td><strong>Classroom Organization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productivity</td>
<td>.054</td>
<td>9.008</td>
<td>7.651</td>
</tr>
<tr>
<td>Instructional learning formats</td>
<td>.051</td>
<td>8.985</td>
<td>7.613</td>
</tr>
<tr>
<td><strong>Instructional Support</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concept development</td>
<td>.052</td>
<td>8.903</td>
<td>7.539</td>
</tr>
<tr>
<td>Quality of feedback</td>
<td>.053</td>
<td>9.092</td>
<td>7.735</td>
</tr>
<tr>
<td>Language modeling</td>
<td>.045</td>
<td>9.188</td>
<td>7.785</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
<th>All-day</th>
<th>Alternating Day</th>
<th>K+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instructional Support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concept development</td>
<td>.025</td>
<td>8.458</td>
<td>9.760</td>
<td>6.444</td>
</tr>
<tr>
<td>Language modeling</td>
<td>.048</td>
<td>8.759</td>
<td>10.132</td>
<td>6.568</td>
</tr>
</tbody>
</table>

*Source: Calculations derived from SPSS – Estimates of Fixed Effects and Estimated Marginal Means*
CHAPTER 5:
CONCLUSIONS, DISCUSSION, and RECOMMENDATIONS

Analysis of Data

This chapter will focus on (1) outcomes and conclusions drawn from research data collected for the current study, (2) discussion of the implications of these conclusions, and (3) recommendations for future research and action based on these findings. First, data collected in this study is compared to that of a number of prior studies utilizing the CLASS. Pianta, La Paro, and Hamre (2008) cited several studies in which the CLASS was the primary instrument used to measure classroom quality. Across studies listed (NCEDL Multi-State Study of Prekindergarten, State-Wide Early Education Programs Study, MyTeachingPartner Study, 4Rs Program Study, Responsive Classroom Approach Study, Induction Study, and the NICHD Study of Early Child Care and Youth Development), Negative Climate tended to show the least variability in scores, while other scores were “adequately distributed across the 7-point scale” (p. 96). Concept Development, Quality of Feedback and Language Modeling tended to have lower scores than the remaining dimensions. This is consistent with findings in the current study, indicating that perhaps the experiences of students in the current study district are similar to those of children across prior studies.

When comparing cited scores (mean, standard deviation and range) from six studies that utilized the CLASS to those recorded for the current study, outcomes were very similar with one interesting difference. Mean scores in the current study, while consistent with prior studies in terms of distribution of scores across dimensions, were
consistently higher than the means of scores for prior studies cited (See Table 5.1 for a comparison). On all dimensions in all studies, the mean scores for dimensions are lower for prior studies than for the current study except for one instance. Mean scores for the K–5 Induction Study were slightly higher than mean scores in the current study on six of the eight dimensions. This may be due to the age of students in the K–5 Induction Study. The authors of the manual suggest that classes in higher grades tend to have higher scores on some dimensions (Pianta et al., 2008).

Differences in mean scores could also be attributed to scorer error or differences in the ways in which teachers tended to interact in the prior study districts. No reason is provided by the authors for differences in the cited studies. However, since the researcher has demonstrated reliability in scoring, within one point of master scorers more than 80% of the time, it might be assumed that differences in the current study are a result of differences in the classroom rather than the scorer’s reliability. In addition, data from and claims about studies cited in the CLASS Manual suggested that different observation procedures did not dramatically affect CLASS scores, so one may infer that any differences in scoring procedures by the investigator in this study did not dramatically change the outcomes on scores for the classrooms observed. These differences in scores will be discussed more thoroughly later in the chapter.
### Table 5.1.

**Descriptive Statistics (M, SD and Range) on CLASS across Six Studies – Comparison to Current Study**

<table>
<thead>
<tr>
<th>Grades (sample)</th>
<th>Preschool (MS and SWEEP)</th>
<th>Preschool (MTP)</th>
<th>Kdg. (MS)</th>
<th>Third Grade (4Rs)</th>
<th>K-5 (Induction)</th>
<th>Grades 1-5 (Responsive Classroom)</th>
<th>Kdg. (Current Study)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M = Mean; SD = Standard Deviation; Kdg. = Kindergarten; COS = Classroom Observation System; N = number of classes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>N</strong></td>
<td>695</td>
<td>164</td>
<td>730</td>
<td>82</td>
<td>22</td>
<td>88</td>
</tr>
<tr>
<td>Positive Climate</td>
<td><strong>M</strong></td>
<td>5.28</td>
<td>5.21</td>
<td>5.14</td>
<td>4.44</td>
<td>5.43</td>
<td>4.91</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.5 – 7.0</td>
<td>2.7 – 7.0</td>
<td>2.5 – 6.8</td>
<td>2.0 – 6.8</td>
<td>3.3 – 7.0</td>
<td>2.5 – 7.0</td>
</tr>
<tr>
<td>Negative Climate</td>
<td><strong>M</strong></td>
<td>1.55</td>
<td>1.63</td>
<td>1.55</td>
<td>2.22</td>
<td>1.17</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.0 – 6.3</td>
<td>1.0 – 4.6</td>
<td>1.0 – 5.3</td>
<td>1.0 – 6.7</td>
<td>1.0 – 2.5</td>
<td>1.0 – 3.5</td>
</tr>
<tr>
<td>Teacher Sensitivity</td>
<td><strong>M</strong></td>
<td>4.70</td>
<td>4.34</td>
<td>4.64</td>
<td>4.60</td>
<td>5.52</td>
<td>4.74</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.5 – 7.0</td>
<td>2.0 – 7.0</td>
<td>1.9 – 6.8</td>
<td>1.3 – 6.8</td>
<td>3.0 – 7.0</td>
<td>2.5 – 6.5</td>
</tr>
<tr>
<td>Risk for Student Perspectives</td>
<td><strong>M</strong></td>
<td>4.36</td>
<td>4.28</td>
<td>4.77</td>
<td>4.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>2.0 – 6.3</td>
<td>1.7 – 6.0</td>
<td>2.8 – 6.5</td>
<td>2.75 – 7.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior Management</td>
<td><strong>M</strong></td>
<td>4.97</td>
<td>4.94</td>
<td>5.18</td>
<td>4.98</td>
<td>5.90</td>
<td>5.14</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.0 – 7.0</td>
<td>2.0 – 6.7</td>
<td>1.9 – 7.0</td>
<td>1.0 – 7.0</td>
<td>4.0 – 7.0</td>
<td>3.5 – 7.0</td>
</tr>
<tr>
<td>Productivity</td>
<td><strong>M</strong></td>
<td>4.5</td>
<td>5.41</td>
<td>4.67</td>
<td>4.69</td>
<td>5.96</td>
<td>4.98</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.0 – 6.9</td>
<td>3.0 – 7.0</td>
<td>1.8 – 6.5</td>
<td>1.5 – 6.8</td>
<td>4.3 – 7.0</td>
<td>2.5 – 7.0</td>
</tr>
<tr>
<td>Instruction</td>
<td><strong>M</strong></td>
<td>3.90</td>
<td>4.57</td>
<td>4.11</td>
<td>4.21</td>
<td>5.22</td>
<td>4.23</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.0 – 7.0</td>
<td>2.0 – 6.3</td>
<td>1.1 – 6.1</td>
<td>1.0 – 6.5</td>
<td>3.3 – 7.0</td>
<td>2.5 – 6.0</td>
</tr>
<tr>
<td>Learning Formats</td>
<td><strong>M</strong></td>
<td>2.09</td>
<td>2.69</td>
<td>2.11</td>
<td>3.84</td>
<td>4.22</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.0 – 5.3</td>
<td>1.5 – 4.4</td>
<td>1.0 – 5.2</td>
<td>1.0 – 6.5</td>
<td>2.0 – 7.0</td>
<td>1.5 – 6.5</td>
</tr>
<tr>
<td>Concept Development</td>
<td><strong>M</strong></td>
<td>2.04</td>
<td>2.87</td>
<td>1.84</td>
<td>3.54</td>
<td>4.61</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1.0 – 5.8</td>
<td>1.0 – 6.0</td>
<td>1.0 – 5.2</td>
<td>1.0 – 6.0</td>
<td>2.0 – 7.0</td>
<td>2.0 – 7.0</td>
</tr>
</tbody>
</table>

**Language**

| **M** | 2.85 |
| **Range** | 1.0 – 5.0 |

**Modeling**

| **M** | (.73) |
| **Range** | 1.5 – 7.0 |

**Key:** MS = Multi-State Study of Preschool; SWEEP = State-Wide Early Education Programs; MTP = MyTeachingPartner; N = number of classes.


Note: Data from the current study has been added to the data published in the manual. Current data is in the far right column.
Research Questions and Hypotheses

This section will focus on conclusions drawn from data analysis beginning with the research questions and null hypotheses for each question.

Research Question 1: Is kindergarten achievement growth in mathematics related to global classroom quality as measured by the CLASS?

Although the null hypothesis was rejected for this question, the researcher found the outcomes for this question inconclusive. While mathematics achievement growth in kindergarten showed significant positive relationship to global classroom quality at the 0.05 level of significance with a slope of .835 during the first growth period (fall to winter), no significant growth during the second growth period (winter to spring) or when overall growth was measured (fall to spring) was observed.

Research Question 2: Is kindergarten achievement growth in reading related to global classroom quality as measured by the CLASS?

Analysis of the relationship between reading achievement growth as measured by the NWEA MAP and global classroom quality as measured by the CLASS showed no significance in any of the growth periods; the null hypothesis was supported in this case. While scores were typically distributed internally (certain dimensions tended to have higher scores, while others tended to have lower scores), these scores did not relate to achievement in reading.

Research Question 3: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in mathematics?

As stated in Chapter Four, the null hypothesis was rejected for this question. Several of the specific quality dimensions showed a significant relationship to
mathematics achievement growth during the first growth period (fall to winter) at the 0.05 level of significance. The following dimensions (slope coefficients) showed a positive relationship to mathematics growth. Regard for Student Perspectives (.702), Productivity (.794), and Instructional Learning Formats (.871). Concept Development (.747), Quality of Feedback (.808), and Language Modeling (.833) demonstrated significance at the 0.01 level of significance. None of the indicators (dimensions) showed a significant relationship to student mathematics achievement during the winter to spring growth period nor for overall growth.

Research Question 4: Which specific indicators (dimensions), as measured by the CLASS, are related to kindergarten achievement growth in reading?

None of the specific indicators of classroom quality, as measured by the CLASS showed a significant relationship to reading growth during any of the growth periods. The null hypotheses were supported in all cases: fall to winter, winter to spring, and overall.

Discussion

While the fall to winter scores in mathematics were in line with results achieved by other researchers, the winter to spring and overall scores did not support the findings in other studies nor were any of the reading growth scores in this study significantly related to classroom quality. This is inconsistent with outcomes for prior research cited by CLASS developers in the CLASS Manual K – 3 (2008). Pianta and his colleagues cited a number of studies in which the CLASS demonstrated predictive validity for students’ academic and social development. Specifically, the authors cite Howes et al. (in press) and Mashburn et al. (in press):

102
The CLASS was designed to assess classroom-level processes that are directly associated with children’s performance. Results from the NCDEL Multi-State Study provide evidence that classroom quality, as assessed by the CLASS, is associated with children’s performance at the end of preschool as well as gains in their performance across the preschool year. In this study, the association between the CLASS and children’s outcomes was assessed after adjusting for a variety of covariates, including maternal education, ethnicity, and gender. (p. 108)

Another study cited in the CLASS Manual, the NICHD SECCYD, (The Eunice Kennedy Shriver National Institute of Child Health and Human Development Study of Early Child Care and Youth Development), showed similar results using a precursor to the CLASS, called the Classroom Observation System (COS). In fact, the developers of the CLASS claim:

> [t]he CLASS and COS have been used to observe more than 4,000 classrooms across the United States and, as such, represent two of the most extensively used observation measures for preschool through elementary classrooms…these are well-validated tools that provide evidence that classrooms that obtain higher scores on these two scoring systems have students who make greater academic and social progress during the school year. (Pianta et al., 2008, p. 91)

Two additional studies of state-funded preschool programs were conducted by the National Center for Early Development and Learning (NCEDL), the NCEDL Multi-State Study of Prekindergarten and the State-Wide Early Education Programs Study (SWEEP). Both of these studies examined how differences in preschool programs relate to social and academic outcomes for children through first grade. These studies included 694
Additional studies cited by Pianta and his co-researchers (2008) include: the MyTeachingPartner Study, which included 164 preschool classrooms in Virginia, conducted by Robert Pianta; the 4Rs Program, which investigated 82 third through fifth grade classrooms located in New York City, conducted by Dr. Lawrence Aber of New York University; the Responsive Classroom Approach study, which collected data from 88 classrooms in “an urban district in the Northeast (p. 94)” conducted by Dr. Sara Rimm Kaufman at the University of Virginia; the Induction Study, which included 33 first and second-grade classrooms in a Southeastern city, conducted by Pianta; and the NICHD Study of Early Childcare and Youth Development study conducted by the NICHD ECCRN (National Institute of Child Health and Human Development Early Child Care Research Network), which was a longitudinal study of 1,364 children from birth through age 15. For all of the studies listed above, score trends stand out as being consistent across classrooms, and predictive validity was consistently established throughout the studies. Why then did mathematics results for kindergarten students in the current study align with these findings during the first growth period, but conflict with these findings during the second growth period? And, why did reading scores fail to show a relationship to classroom quality in any of the growth periods? There may be a number of reasonable answers to this question. The researcher discusses a few of these in the following section.

As stated previously, observations in the current study were significantly related to achievement scores in mathematics during the first growth period, but not in the second growth period or for overall mathematic academic growth. In addition,
observations in the current study yielded consistently higher mean scores than did classes observed in prior studies cited in the CLASS Manual. If these differences are not due to scorer errors, as may be suggested by the internal relationship of scores (Dimensions that typically yield high scores did so in this study, while dimensions that typically yield low scores also followed the expected pattern.), then perhaps other factors impacted outcomes.

The researcher suggests three possible reasons for these outcomes. First, perhaps district-wide training required for all kindergarten teachers reduced variability in classroom-level characteristics of quality. All teachers participated in HET (Highly Effective Teaching) training as well as literacy training, which proscribed a specific curriculum and procedures for reading. If this training provided more consistency among classroom processes, the similarity among classrooms may have made it more difficult for relationships between classroom quality and student achievement to stand out.

Although there was variability in CLASS scores across classes, from a low of 1.5 to a high of 7.0, the majority of scores fell within the medium range and the researcher noted specific factors which affected the low scores in each case where they occurred. For example, low scores in some instances were related to the activity observed. When a teacher was engaged in assessment, for example, little instructional support was evident. Besides instances such as this, major variances were rare.

Pianta and La Paro (2003) provide support for the above theory in an article reporting longitudinal research involving the CLASS research in more than 2,000 early education classrooms from prekindergarten through first grade. They stated, “The first conclusion we drew from this work was that early education classrooms vary widely in
the activities in which children participate and the quality of the classroom environment…[t]his picture was consistent from prekindergarten through 1st grade” (pp. 27-28). If typical classrooms are highly variable, the lower variability in this study may be reasonably associated with a common characteristic among the classrooms, training.

Related to this, training may have created measurable differences in quality when comparing the current research classrooms to those studied in prior research. This could provide an explanation for the current mean scores exceeding those of prior studies. A third explanation for the differences in outcomes seems less likely – that classroom quality simply does not relate to student achievement in this district.

Pianta and La Paro (2003) noted another significant trend in early education classrooms (defined as prekindergarten through first grade). They said, “[d]espite the exceptional variability in activities, an overall picture emerged”, which can be characterized as “socially positive but instructionally passive” (pp. 27-28). While classes in the current study may not be labeled “instructionally passive”, there was some variability among scores depending upon the domain. The mean score (standard deviation) for Emotional Support was 5.51 (1.18) and the mean score for Classroom Organization was 5.54 (1.01). However, the mean score for Instructional Support was only 4.15 (1.55), almost a full point lower on the Likert-type scale after considering the standard deviation. So, while Instructional Support was not within the low range, it was clearly lower than Emotional Support and Classroom Organization.

The Effect of HET (Highly Effective Teaching) Model

As mentioned above, one condition which may have affected outcomes in this study is the widespread implementation of the Highly Effective Teaching Model (HET)
in the study district. This model, created by Susan Kovalik, has been used in the study district since 2003. It has at its root brain research and empirically tested strategies for creating effective learning environments in the classroom. In addition to a focus on the biology of learning, this model trains teachers to use instructional strategies which have been shown to improve learner engagement and outcomes. The HET model is based on five principles drawn from brain research. They are:

1. Intelligence is a function of experience.
2. Learning is an inseparable partnership between the brain and the body.
   a) Emotion is the gatekeeper to learning and performance.
   b) Movement enhances learning.
3. There are multiple intelligences or ways of solving problems and/or producing products.
4. Learning is a two-step process.
   a) Step one: Making meaning through pattern seeking
   b) Step two: Developing a mental program for using what we understand and wiring it into long-term memory
5. Personality impacts learning and performance. (Kovalik & Olsen, 2009, p. xi)

In addition to the five principles listed above that were taken from brain research, Kovalik and Olsen (2009) suggest that there are nine body/brain compatible elements of curriculum, which are “the primary ways of translating brain research into action in the classroom (p. xiv).” These are: absence of threat/nurturing reflective thinking.
meaningful content, movement, enriched environment, choices, adequate time, collaboration, immediate feedback, and mastery/application.

All kindergarten teachers in the district have been trained and are expected to implement these practices into their classrooms. So, there is a high level of consistency among classes on elements identified as “highly effective” according to the model. This is relevant because the dimensions measured by the CLASS have a strong correlation to the principles promoted in the HET model. If all teachers have been trained to teach using similar strategies, are expected to use the same curriculum, and must create specific environments in their classrooms based on this model, it may be reasonable to assume that teachers in this district may demonstrate less variability on dimensions of classroom quality than teachers who were not trained under this model. It may also provide a rationale for the higher mean scores attained by teachers in the current research group.

*Lifelong Guidelines and LIFESKILLS*

According to the Susan Kovalik (2009),

Lifelong Guidelines and LIFESKILLS provide the basis for three important functions of a bodybrain-compatible learning environment:

- They are the agreed-upon behaviors for all and thus replace the rules of traditional ‘discipline’ programs.
- They provide a safe environment for creating and maintaining a sense of community.
- They describe the behaviors of civil discourse which are the foundation of citizenship. (p. 9.1)
The Lifelong Guidelines and LIFESKILLS are another integral part of the routine in all district elementary schools included in the study (See Table 5.2 for a complete listing of Lifelong Guidelines and LIFESKILLS). The procedures and processes inspired by these guidelines for living together in community are consistently taught and reinforced in every classroom across the district. As a matter of policy, these guidelines are posted and taught to students in every classroom, and procedures for how these are to be implemented are part of each teacher’s classroom management plan (Assistant Superintendent of Instruction, personal communication, March 9, 2010). In addition, emphasis on creation of a non-threatening environment, modeling of positive behavior, attention to Multiple Intelligence Theory, active learning, and integration of learning across the content areas are essential program elements taken from HET (Kovalik & Olsen, 2009). Each of these supports the creation of a positive, high quality classroom environment as measured by the CLASS. Perhaps implementation of these skills and guidelines in the study district has also contributed to higher quality in classrooms across the district than that found in studies cited in the testing manual.

Perhaps prior training and the similarity among classrooms in the study district may have impacted outcomes on the CLASS. In addition, it is possible that these same similarities may have become more prevalent over time. The longer students were in these classrooms, the more their experiences became similar. If this is true, it may (at least in part) explain why the relationship between CLASS scores in mathematics was statistically significant early in the year, but diminished with time and additional instruction.
Table 5.2
Lifelong Guidelines and LIFESKILLS

<table>
<thead>
<tr>
<th>Lifelong Guidelines</th>
<th>Lifeskills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trustworthiness: to act in a manner that makes one worthy of trust and confidence</td>
<td>Caring: To feel and show concern for others</td>
</tr>
<tr>
<td>Truthfulness: To act with personal responsibility and mental accountability</td>
<td>Common Sense: To use good judgment</td>
</tr>
<tr>
<td>Active Listening: To listen attentively and with the intention of understanding</td>
<td>Cooperation: To work together toward a common goal or purpose</td>
</tr>
<tr>
<td>No Put-downs: To never use words, actions, and/or body language that degrade, humiliate, or dishonor others</td>
<td>Courage: To act according to one’s beliefs despite fear of adverse consequences</td>
</tr>
<tr>
<td>Personal Best: To do one’s best given the circumstances and available resources</td>
<td>Curiosity: A desire to investigate and seek understanding of one’s world</td>
</tr>
<tr>
<td></td>
<td>Effort: To do your best</td>
</tr>
<tr>
<td></td>
<td>Flexibility: To be willing to alter plans when necessary</td>
</tr>
<tr>
<td></td>
<td>Friendships: To make and keep a friend through mutual trust and understanding</td>
</tr>
<tr>
<td></td>
<td>Initiative: To do something of one’s own free will, because it needs to be done</td>
</tr>
<tr>
<td></td>
<td>Integrity: To act according to a sense of what’s right and wrong</td>
</tr>
<tr>
<td></td>
<td>Organization: To plan, arrange, and implement in an orderly way, to keep things orderly and ready to use</td>
</tr>
<tr>
<td></td>
<td>Patience: To wait calmly for someone or something</td>
</tr>
<tr>
<td></td>
<td>Perseverance: To keep at it</td>
</tr>
<tr>
<td></td>
<td>Pride: Satisfaction from doing one’s personal best</td>
</tr>
<tr>
<td></td>
<td>Problem Solving: To create solutions to difficult situations and everyday problems</td>
</tr>
<tr>
<td></td>
<td>Resourcefulness: To respond to challenges and opportunities in innovative and creative ways</td>
</tr>
<tr>
<td></td>
<td>Responsibility: To respond when appropriate; to be accountable for one’s actions</td>
</tr>
<tr>
<td></td>
<td>Sense of Humor: To laugh and be playful without harming others</td>
</tr>
</tbody>
</table>

Limitations of the Study

Number of variables. One of the limitations posed by this study is the sheer number of variables addressed in the analysis. The larger the number of variables and statistical analyses applied, the greater the likelihood that a Type 1 Error will occur. This happens when “we believe that there is a genuine effect in our population, when in fact there isn’t” (Field, 2005, p.31). Because children were tested three times during the study, there were three sets of scores for both mathematics and reading. So comparison of fall to winter, winter to spring, and overall academic growth included six different sets of growth scores. In addition, the CLASS is organized into three broad domains of classroom quality: Emotional Support, Classroom Management, and Instructional Support. Within these domains, ten more specific dimensions of classroom quality were assessed. To determine the relationship among all of these variables required over 70 different analyses with multiple covariates for each. This may have elevated the Type 1 Error rating to almost 1.0.

Because of the above, one must be cautious about generalizing the findings of this study to other groups. Perhaps it is reasonable to assume that similar outcomes are possible in a similar school district with similar students who are also taught in schools that implement HET, but again, one should do so with caution. The elevated Type 1 Error rating makes it impossible to determine which of the significant outcomes were actually significant and which were the results of random chance.

Program variations. Another limitation of the current study is the program variability found among district kindergarten classrooms. While some children attended all-day every-day kindergarten, others attended alternating days or were part of the K+
program, which was implemented in two different formats. Because of these variations, it is difficult to be certain if outcomes were related to classroom quality differences or program differences. A number of studies provide evidence that length of day may impact student achievement, so perhaps these differences affected outcomes in the current study (Baskett et al., 2005; Clark & Kirk, 2000; Gullo, 2000; Lee et al., 2006; Nunnely, 1996; Plucker et al., 2004; Rothenburg, 1995; Walston & West, 2004; Wood, January 29, 2004; & Zvoch et al., 2008). While program was controlled for, an additional confound may have been related. Program assignments were made based upon prior achievement and socio-economic status (identified by free, reduced or full-pay meal status). This factor may have interacted in ways not identified through analysis.

Class-size variation. Research also suggests that class size may have an impact on learner outcomes (Barnett et al., 2007; Bryant et al., 2005a; Denton, 2001; Finn & Pannozzo, 2004; Haenn, 2002; Mead, 2008; NICHD, 2002; Yan & Lin, 2005; & Zahorik, 1999). Class size which varied significantly in a few situations was not controlled for in analysis. There were some classes in which there were as few as a dozen children and others had enrollments of up to 24.

Socio-economic differences. In addition (or in relation) to the above limitations encountered in the study, socio-economic differences may have had an impact on outcomes. Although this was considered in the analysis by controlling for socio-economic status through meal payment designation (free, reduced or full-pay), students were placed in classrooms and/or programs based on their Title I status. This means that children in all-day every-day and K+ classrooms tended to come from homes with lower incomes than did children in the alternating day programs. Again, the researcher
controlled for this effect, but program type was shown to be significantly related in a few cases, with children in alternating day programs making more academic progress. It is difficult to know if this is due to classroom quality differences or the demographics of the student populations within each program model or class. In any case, it seems counter-intuitive that children who received less instruction should make greater progress than those who received more instruction. Perhaps the relationship between income levels in the home and program model should come into consideration.

Reliability and validity of standardized tests. Ann Benjamin of the University of Massachusetts, Lowell may have said it best. She said, “Most teachers know intuitively that testing children under the age of seven or eight yields questionable results. Nonetheless, the ‘testing craze’ (Why the Testing Craze”, 1999) is likely to continue” (Cited in Brewer, 2007, p. 218). This may simply be because no better means of assessing and comparing children’s academic progress has been identified, but in any case, it is a reality under which this researcher and others must operate. This recognized, a final limitation of the current study may be related to the developmental stage of the students involved and the form of assessment utilized to measure student achievement.

Although the NWEA MAP is specifically designed for young children and includes supports to help children succeed (auditory instructions, picture format, practice tests, and simple click applications) (NWEA MAP Features, 2008; NWEA MAP Overview, 2008;), research suggests that standardized tests may not be reliable indicators of student achievement because outcomes may be skewed by children’s lack of experience with computers, poor eye-hand coordination, or inability to focus for extended periods of time. Woodfield and Lewis (2003) “found teachers unhappy with traditional
state-mandated tests because these tests didn’t truly measure, report or track student growth…” (para 4). NWEA MAP literature suggests that these concerns have been overcome with this assessment; however, teachers in the study district expressed concern over whether the test would really measure academic growth for all students. So, there is a possibility that early scores were not valid (measuring what was intended) or reliable (consistent, dependable, or repeatable) due to the developmental level of subjects.

Calculation of global quality score. To calculate a global quality score for each class, the researcher calculated a mean score for each dimension of quality assessed using the CLASS. The means for these ten dimension scores were then added together, and divided by ten to achieve an overall (or global) mean score for classroom quality. Although this was an extension of the method used to calculate composite domain scores, there is no prior research indicating calculation in this manner, the findings may not truly represent global quality in the way intended.

Recommendations for Future Research and Action

Replication of the Study in the Current District

Changes in the program format in the current study district may make a difference in outcomes on the CLASS, as well as upon how those scores relate to student achievement in mathematics and reading. During the study period, three different kindergarten program models were implemented across the district. While some children attended (1) all-day every-day kindergarten, other children attended (2) an alternating day program (either all day Monday and Wednesday with a half-day on Friday morning or all day Tuesday and Thursday with a half day on Friday afternoon), or (3) K+, which could have taken two different formats. Children enrolled in K+ classes may have gone to
school all-day every-day with the same teacher, or they may have gone on alternating
days to their class of record and on the remaining days attended in a classroom with a
different (K+) teacher.

The variability in programming may be a cause for concern that outcomes could
not be reasonably compared. Children “at risk” were clustered in K+ or all-day every-day
classes. These children were identified by free or reduced meal status or low achievement
on pre-enrollment screening. However, children from higher socioeconomic backgrounds
(determined by full-pay meal status) and those who scored well on screening tests prior to
placement were placed in alternating-day programs. These differences in classroom
makeup may be reflected in the higher scores for children attending alternating day
programs. Even though they received fewer hours of instruction, they tended to make
greater achievement gains. Perhaps this is due to higher ability of the children placed in
alternating-day programs, more enriched environments in children’s homes, or some
other factor not identified. If classes were evenly or randomly distributed in terms of
prior student achievement and socioeconomic status, these outcome differences by
program may not have been evident.

The researcher recommends a future replication study in the same school district
because all kindergarten classes in the district are now all-day every-day. This consistent
program format would allow one variable (program) to be removed from the formula,
reducing the likelihood of a Type I Error.

Replication of the Study in Another HET District

The researcher recommends that a follow-up study be executed in another school
district in which the Highly Effective Teaching model has been implemented to
determine if the strategies do, after all, increase classroom quality across all dimensions and domains. This may be a valuable undertaking. If implementation of the HET model can be shown to improve classroom quality and this improvement can be correlated to higher achievement, the value of the HET model may be supported. However, if a significant relationship cannot be determined, one might assume differences were unrelated to implementation of HET in the study district.

**Continued Training**

According to Fukkink and Lont (2007), “[c]orrelational research has suggested that the training of caregivers is a cornerstone for quality in early care” (p. 294). If this is also true at the kindergarten level, and research suggests it is (Barnett, 2004; Gerber et al., 2007; Kelly & Camilli, 2007), it would seem logical that continued training, particularly in programs that have shown a history of increased achievement and quality, would result in higher quality classrooms and greater achievement for kindergarten children across the district. The researcher recommends that the local district continue to provide training in the Highly Effective Teaching model and that perhaps, the addition of CLASS training would enhance classroom quality and achievement to an even greater extent than HET alone.

If CLASS scores are indeed predictive of achievement (Pianta et al., 2008) then training teachers to improve classroom quality through the dimensions identified in the CLASS may be a means of increasing academic outcomes for children across the district, leading to greater numbers of schools meeting annual yearly progress goals. According to Pianta (2005/2006), using the CLASS for teacher observations provides a “standard way of noting teachers’ strengths and weaknesses and forms the basis from which professional
development can support teachers’ high quality implementation and improve teacher-child interactions” (para. 6).

Summary and Conclusions

School districts across the nation are dealing with pressure from the federal government and states to meet NCLB mandates “to achieve unprecedented educational progress…ensuring that all students—and all subgroups—meet the state’s proficiency goals” (Sunderman & Orfield, 2007, p. 138). Teachers strive to address these concerns by preparing students for future grades as well as standardized testing. It is a delicate balancing act, juggling academics with developmentally appropriate practice. To address these issues, classroom observations and analyses were conducted with the following goal in mind. Ten dimensions of classroom quality were assessed using a standardized assessment instrument (CLASS). Student academic achievement in mathematics and reading as measured by NWEA MAP Primary Version was compared with global classroom quality ratings and specific quality indicators to determine if classroom quality and student achievement in mathematics and reading were related. If relationships could be established, perhaps specific training to help teachers improve classroom quality, and in turn, student achievement could be implemented resulting in schools meeting annual yearly progress goals and NCLB mandates.

While the answer to the above question yielded inconclusive results it seems that additional research may support continued use of the CLASS to assess classroom quality. In addition to assessment using the CLASS, perhaps training teachers to identify and implement practices consistent with a high quality classroom may translate into higher achievement for all
References


in state-funded pre-kindergarten programs and associations with teacher, program and classroom characteristics. *Early Childhood Research Quarterly, 22*(1), 3-17.


135

Sunderman, G. & Orfield, G. (2007, October). Do states have the capacity to meet the NCLB mandates? *Phi Delta Kappan, 89*(2), 137-139.


Yan, W. & Lin, Q. (2005). Effects of class size and length of day on kindergartners’ academic achievement: findings from Early Childhood Longitudinal Study. *Early Education and Development*, 16(1), 49-68.

Appendix A

Please complete the following by writing a response or marking all choices that apply.

Teacher Name:  School:  Schedule:  ADED  MWF  TRF  EXT  K+
Degree:  Bach  Bach Plus  Master’s  Master's Plus  Specialist  Other:
# Years Teaching (including current):  # Years Teaching Kdg.
# Students:  # Assistants:  Asst. Hours Daily:  # Other Adults:  Hours Daily:
# Special Needs  # Speech

Any special circumstances of which I should be aware:

<table>
<thead>
<tr>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B
SUMMARY OF CLASS DOMAINS AND DIMENSIONS
SUMMARY OF CLASS DOMAINS AND DIMENSIONS

I. EMOTIONAL SUPPORT (DOMAIN)

a. Positive Climate (Dimension – indicators and specific behavior markers are listed below)
   i. Relationships
      1. Physical proximity
      2. Shared activities
      3. Peer assistance
      4. Matched affect
   ii. Positive Affect
      1. Smiling
      2. Laughter
      3. Enthusiasm
   iii. Positive Communication
      1. Verbal affection
      2. Physical affection
      3. Positive expectations
   iv. Respect
      1. Eye contact
      2. Warm, calm voice
      3. Respectful language
      4. Cooperation and sharing

b. Negative Climate (Dimension – indicators and specific behavior markers are listed below)
   i. Negative Affect
      1. Irritability
      2. Anger
      3. Harsh voice
      4. Peer aggression
      5. Disconnected or escalating negativity
   ii. Punitive Control
      1. Yelling
      2. Threats
      3. Physical control
      4. Harsh punishment
   iii. Sarcasm/Disrespect
      1. Sarcastic voice/statement
      2. Teasing
      3. Humiliation
   iv. Severe Negativity
      1. Victimization
      2. Bullying
      3. Physical punishment
c. Teacher Sensitivity (Dimension – indicators and specific behavior markers are listed below)
   i. Awareness
      1. Anticipates problems and plans appropriately
      2. Notices lack of understanding and/or difficulties
   ii. Responsiveness
      1. Acknowledges emotions
      2. Provides comfort and assistance
      3. Provides individual support
   iii. Addresses Problems
      1. Helps in an effective and timely manner
      2. Helps resolve problems
   iv. Student Comfort
      1. Seeks comfort and guidance
      2. Freely participates
      3. Takes risks

d. Regard for Student Perspectives (Dimension – indicators are listed below)
   1. Flexibility and Student Focus
   2. Shows flexibility
   3. Incorporates student’s ideas
   4. Follows lead
   ii. Support for Autonomy and Leadership
      1. Allows choice
      2. Allows students to lead lessons
      3. Gives students responsibilities
   iii. Student Expression
      1. Encourages student talk
      2. Elicits ideas and/or perspectives
   iv. Restriction of Movement
      1. Allows movement
      2. Is not rigid

II. CLASSROOM ORGANIZATION – (DOMAIN)
a. Behavior Management (Dimension – indicators and specific behavior markers are listed below)
   i. Clear Behavior Expectations
      1. Clear expectations
      2. Consistency
      3. Clarity of rules
   ii. Proactive
      1. Anticipates problem behavior or escalation
      2. Low reactivity
      3. Monitors
   iii. Redirection of Misbehavior
1. Effective reduction of misbehavior
2. Attention to the positive
3. Uses subtle cues to redirect
4. Efficient redirection

iv. Student Behavior
1. Frequent compliance
2. Little aggression and defiance

b. Productivity
i. Maximizing Learning Time
   1. Provision of activities
   2. Choice when finished
   3. Few disruptions
   4. Effective completion of managerial tasks
   5. Pacing

ii. Routines
   1. Students know what to do
   2. Clear instructions
   3. Little wandering

iii. Transitions
   1. Brief
   2. Explicit follow-through
   3. Learning opportunities within

iv. Preparation
   1. Materials ready and accessible
   2. Knows lessons

c. Instructional Learning Format
i. Effective Facilitation
   1. Teacher involvement
   2. Effective questioning
   3. Expands student involvement

ii. Variety of Modalities and Materials
   1. Range of auditory, visual, and movement opportunities
   2. Interesting & creative materials
   3. Hands-on opportunities

iii. Student Interest
   1. Active participation
   2. Listening
   3. Focused attention

iv. Clarity of Learning Objectives
   1. Advanced organizers
   2. Summaries
   3. Reorientation statements

III. INSTRUCTIONAL SUPPORT (DOMAIN)
a. Concept Development
   i. Analysis and Reasoning
1. Why and/or how questions
2. Problem solving
3. Prediction/experimentation
4. Classification/comparison
5. Evaluation

ii. Creating
   1. Brainstorming
   2. Planning
   3. producing

iii. Integration
   1. Connect concepts
   2. Integrates with previous knowledge

iv. Connections to the Real World
   1. Real-world applications
   2. Related to students’ lives

b. Quality of Feedback
   i. Scaffolding
      1. Hints
      2. Assistance
   ii. Feedback Loops
      1. Back-and-forth exchanges
      2. Persistence by teacher
      3. Follow-up questions
   iii. Prompting Thought Processes
      1. Asks students to explain thinking
      2. Queries responses and actions
   iv. Providing Information
      1. Expansion
      2. Clarification
      3. Specific feedback
   v. Encouragement and Affirmation
      1. Recognition
      2. Reinforcement
      3. Student persistence

c. Language Modeling
   i. Frequent Conversations
      1. Back-and-forth exchanges
      2. Contingent responding
      3. Peer conversations
   ii. Open-ended Questions
      1. Questions require more than a one-word response
      2. Students respond
   iii. Repetition and Extension
      1. Repeats
      2. Extends/elaborates
iv. Self- and Parallel Talk
   1. Maps own actions with language
   2. Maps student action with language
v. Advanced Language
   1. Variety of words
   2. Connected to familiar words and/or ideas (Pinata et al., 2008)