Abstract

Title of dissertation: EVALUATING WHICH CLASSROOM AND STUDENT VARIABLES IN AN EARLY CHILDHOOD PROGRAM BEST PREDICT STUDENT LANGUAGE AND LITERACY ACHIEVEMENT

Lydia Janeva Carlis, Doctor of Philosophy, 2011

Dissertation directed by: Professor Rebecca Silverman
Department of Special Education

Sufficient evidence exists that early intervention for students at-risk of school failure may prevent reading difficulties. In addition, research has identified several teacher, classroom, and student variables that correlate with students' academic progress in language and literacy domains. This research aimed to expand on existing research by analyzing the relationships between teacher and peer characteristics and language and literacy achievement, and change in achievement, for 431 three and four-year old children attending three Early Reading First funded public charter school programs in 29 Washington, DC classrooms. Four research questions were posed to answer the following: What are the peer and teacher variables that predict achievement, or change in achievement, on norm- or criterion-referenced language and literacy measures for children enrolled in a federally supported universal preschool program? Two-level hierarchical linear modeling (HLM) was conducted with students nested in classrooms. Results from the HLM indicate that peer and teacher characteristics helped predict three and four-year olds' achievement on nationally normed standardized language and literacy

assessments. Peer and teacher characteristics also predicted achievement and change in achievement on curriculum-based measures of language and literacy development. These findings expand the research on teacher and peer characteristics predictive of student language and literacy achievement. Implications from these findings, strengths and limitations of this dissertation research, and future research directions are discussed.

EVALUATING WHICH CLASSROOM AND STUDENT VARIABLES IN AN EARLY CHILDHOOD PROGRAM BEST PREDICT STUDENT LANGUAGE AND LITERACY ACHIEVEMENT

by

Lydia Janeva Carlis

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Rebecca Silverman, Chair
Professor Robert Croninger
Professor Joan Lieber
Professor Margaret McLaughlin
Professor Deborah L. Speece

Table of Contents

LIST OF TABLES ..... **iii**

LIST OF FIGURES ..... **v**

CHAPTER 1: INTRODUCTION **1**

STATEMENT OF THE PROBLEM 1
ASSESSING YOUNG CHILDREN'S LANGUAGE AND LITERACY DEVELOPMENT 4
CONCEPTUAL FRAMEWORK FOR CURRENT LITERATURE REVIEW 6
OVERVIEW OF RESEARCH ON ECOLOGICAL FACTORS 8
RESEARCH QUESTIONS 13
DEFINITIONS OF KEY TERMS 15

CHAPTER 2: REVIEW OF THE LITERATURE **17**

METHOD 17
CONTENT ANALYSIS 19
METHODOLOGICAL ANALYSIS 41
SUMMARY OF CONTENT AND METHODOLOGY REVIEWS 49
DISCUSSION AND RATIONALE FOR STUDY 51

CHAPTER 3: METHODOLOGY **54**

RESEARCH DESIGN 54
RESEARCH QUESTIONS AND HYPOTHESES 55
SAMPLE 60
VARIABLES 64
PROCEDURES 75
DATA ANALYSES 82

CHAPTER 4: RESULTS **89**

PRINCIPAL COMPONENTS ANALYSIS 90
DATA EXPLORATION 94
MULTIPLE REGRESSION (MR) ANALYSIS 101
HIERARCHICAL LINEAR MODELING (HLM) 110
SUMMARY OF ANALYSES 125

CHAPTER 5: DISCUSSION & CONCLUSIONS **131**

FINDINGS FROM THE CURRENT STUDY 134
POTENTIAL IMPLICATIONS 153
LIMITATIONS 157
CONCLUSIONS 158

REFERENCES **172**

List of Tables

Table 1.1 Seminal Research Comparison…… 160

Table 2.1 Content Summary of Studies…… …… 162

Table 2.2 Ecological Factors Considered by Study……..... 166

Table 2.3 Descriptive Data by Study…… … 167

Table 2.4 Mean Effect Sizes for Dependent Variables Across Literature Reviews ……..168

Table 2.5 Definitions for Internal, Construct, Statistical Conclusion, and External Validity Criteria…… …… 170

Table 3.1 2009-10 Student Demographics for DCPS, DCPEL and the Analytic Sample. . . . ……. 64

Table 3.2 Names and Descriptions for Variables Used in Study. ……..... 66

Table 3.3 Student Assessment and Progress Monitoring System…… 68

Table 3.4 CLASS Domains and Dimensions……... 71

Table 3.5 ELLCO Domains and Items . . . . .. …….. 72

Table 4.1 Analytic Sample 2009-10 Averages for Dependent Variables ….. 91

Table 4.2 Factor Loadings for Principal Components Analysis of DCPEL 09-10 Student Achievement Data. . . ……..... 93

Table 4.3 2009-10 Dependent Variable Averages for DCPEL and the Analytic Sample. . . ……..... . . . . ……95

Table 4.4 Descriptive data exploring differences in performance on norm- and criterion-referenced assessments by DCPEL children whose families paid or received free or reduced price lunch during the 2009-10 school year. .96

Table 4.5 Correlations and variance inflation factors between dependent variables in the analytic sample. . . . . ……... . 97

Table 4.6 DCPEL 2009-10 Averages Across CLASS Domains. . . . . ….... 98

Table 4.7 DCPEL 2009-10 Averages Across ELLCO Domains and Subscales… . 98

Table 4.8 Descriptive data exploring differences in average instructional quality experienced by DCPEL children whose families paid or received free or reduced price lunch during the 2009-10 school year. . . . …….. 99

Table 4.9 Correlations and variance inflation factors between classroom peer, instruction and quality variables in the analytic sample. . . …. …….. .. 100

Table 4.10 Summary of Multiple Regression Analysis for Variables Predicting Outcome Language (OUT_LAN). . . . .. ……104

Table 4.11 Summary of Multiple Regression Analysis for Variables Predicting Outcome Literacy (OUT_LIT). . . . . ……... . . .. .104

Table 4.12 Summary of Multiple Regression Analysis for Variables Predicting Spring Language (SP_LAN). . . . . .. ……108

Table 4.13 Summary of Multiple Regression Analysis for Variables Predicting Fall-to-Spring Language Change (FS_LAN)……. 108

Table 4.14 Summary of Multiple Regression Analysis for Variables Predicting Spring Literacy (SP_LIT). . . . . ……... . . . 108

Table 4.15 Summary of Multiple Regression Analysis for Variables Predicting Fall-to-Spring Literacy Change (FS_LIT). . …… 108

Table 4.16 Level 1 unconditional model for student language and literacy achievement in SY2009-10. . . . . ……... 111

Table 4.17 Level 1 conditional model for student language and literacy achievement in SY2009-10. ……115

Table 4.18 Level 2 fully conditional model estimating outcome language and literacy achievement for the 2009-10 school year. . . . ……... 118

Table 4.19 Level 2 fully conditional model estimating progress monitoring language and literacy achievement and change for the 2009-10 school year.. 124

Table 4.20 Estimates for proportion of variance explained by level 2 hierarchical linear models for the dependent variables. . . . …….. 125

Table 4.21 Coefficients for predictor variables found significant across the study dependent variables. . . . ……. 130

List of Figures

Figure 1 Adults and environment as mediators of children's understandings in early language and literacy activities. . . . . ……... . . .. .161

Figure 2 Multilevel model for evaluating the effects of child and classroom characteristics on children's learning in language and literacy……. . .. 171

Evaluating Which Classroom and Student Variables in an Early Childhood Program Best Predict Student Language and Literacy Achievement

Chapter 1: Introduction

Statement of the Problem

Reading failure is prevalent among children from low socioeconomic backgrounds, children who are members of racial minority groups, and children whose native language is not English (National Research Council [NRC], 1999). Whitehurst and Lonigan (1998, 2001) suggest that literacy skills as early as the beginning of kindergarten are predictive of later academic achievement, and children who arrive at kindergarten with language and literacy deficits are at highest risk of experiencing difficulty in school. Research consistently demonstrates that early intervention can prevent or remediate students' academic difficulties (Lennon & Slesinski, 1999; Torgesen, 2002; Vellutino, Scanlon, & Tanzman, 1998).

A major concern for researchers, policymakers and practitioners alike is the limited language and literacy skills with which many students enter school. A 2002 study by the National Center for Education Statistics (NCES) suggests that minority students and English Language Learners (ELL) enter preschool at a significant disadvantage in letter identification, phonemic awareness, and print awareness proficiencies, known to be predictors of later reading success. In addition, Hart and Risley (1995) identified early language ability as an area of importance, with children from impoverished backgrounds beginning school with many fewer words than their more advantaged peers. Furthermore, Scarborough's 1998 meta-analysis of 61 studies of kindergarten literacy interventions found that while only 5 - 10% of children who start out reading well

experience problems later on in school, 65 – 75% of children who encounter problems in kindergarten experience difficulties throughout their school careers, putting them at-risk for dropping out of school, criminal activity, and unemployment (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Masse & Barnett, 2003). Thus, researchers and policy makers have begun to study the effectiveness of public pre-kindergarten in preventing subsequent school failure and "closing the achievement gap" for students at risk of school failure (Barnett, Lamy, & Jung, 2005; Christina & Nicholson-Goodman, 2005; Perez-Johnson & Maynard, 2007).

The United States government has been working to close the achievement gap for many years. From the inception of government-funded early childhood care programs, focus has centered on children of impoverished backgrounds. Researchers have also shown interest in examining the effects of early childhood education (ECE) on later academic and social success. Three seminal experimental research projects are consistently cited to support the effectiveness of early intervention and education, and to provide rationale for investment of local, state and federal dollars into early childhood education: (a) High Scope/Perry Preschool Project (Weikart, Deloria, Lawser, & Wiegerink, 1970); (b) Carolina Abecedarian Project (Ramey & Campbell, 1984); and (c) Chicago Parent-Child Center (CPC; Reynolds, Temple, Robertson, & Mann, 2001a, 2001b).

Each of the three studies controlled for race/ethnicity, SES, teacher to student ratio, teacher education, length of school day and year, curriculum, and family involvement (Table 1). Longitudinal benefits of program participation include social, academic, cognitive and financial advantages compared to control group peers (RAND

Corporation, n.d.), demonstrating that early intervention can have a significant impact on children's long-term achievement. However, two of the studies had limited racial/ethnic and socioeconomic diversity. Although 2% of children in the Abecedarian study were Caucasian, and 5% of CPC participants were Hispanic, participants across studies were from low SES backgrounds and primarily African American.

These longitudinal research studies on early intervention have yielded dramatic positive results across cognitive, academic, and social domains (Belfield, Nores, Barnett & Schweinhart, 2006; Campbell, Pungello, Miller-Johnson,Burchinal & Ramey, 2001; Reynolds, Ou & Topitzes, 2004). Children in the seminal studies were all identified with at least two commonly identified risk factors (i.e., minority status [African American] and low SES; NRC, 1999). Nonetheless, it is unclear if the effects found in these highly controlled experiments can be generalized across typical public preschool schools and classrooms, which programmatic factors most impact student achievement, or if benefits extend to students across all of the previously identified risk categories.

Additionally, changes in the program characteristics for publicly funded early childhood programs since these programs began may also alter the potential effects of early childhood education as an intervention. Examples of these changes, experienced differentially by US states, are the rapid growth of second language learners in schools, the provision of school choice to families through public charter programs, and the introduction of universal pre-kindergarten programs that provide services to all children regardless of SES (Barnett, Friedman, Hustdet, Stevenson-Boyd, 2009). Thus, although these studies offer support for the effectiveness of early intervention for the achievement of African American students from low SES backgrounds, it is not clear that these effects

can be replicated in today's traditional public or public charter school classrooms with increasingly diverse populations. Analysis of outcomes for children from a variety of racial, ethnic and economic backgrounds in less controlled public school settings will expand the research on characteristics associated with young children's academic achievement. The focus of the current study is the effect of child, peer and teacher variables on students' language and literacy development in public early education programs serving at-risk students, using norm-referenced standardized outcome assessments and criterion-referenced progress monitoring assessments to measure student achievement.

Assessing Young Children's Language and Literacy Development

Young children's language and literacy achievement is a known predictor of later academic achievement (Whitehurst & Lonigan, 1998; 2001). In the current study, language is defined as the individual and connected vocabulary a child can comprehend and communicate orally, through listening and speaking (Dunn & Dunn, 2007; Lonigan, Wagner, Torgesen & Rashotte, 2000; Williams, 2007). Literacy is defined here as pre-reading skills needed by children to become able readers. These pre-reading skills include concepts about print, alphabet knowledge, letter-sound relationships, print and book knowledge, emergent writing, and phonological and phonemic awareness (Lonigan et al., 2000; McBridge-Chang, 1998; Whitehurst, 2001; Whitehurst & Lonigan, 2001).

Educators must identify and use valid language and literacy assessment options for all three subgroups identified in 1999 by the NRC (i.e., low SES, ethnic minority, and second language learners). According to the Institute of Education Sciences, English language learners (ELLs) can be screened on the same early reading indicators as native

English language speakers, including phonological awareness, letter knowledge, and word and text reading (Gersten et al., 2007). Additionally, ELLs can and should be tested as regularly as their native English-speaking peers (Gersten et al., 2007). Thus, similar targeted assessment in early reading indicators can be effective for all three of the earlier identified at-risk populations.

Furthermore, it is important that educators use current assessment data to make appropriate differentiated instructional decisions. One core principle of effective differentiated instruction is recognizing children's strengths and needs through systematic screening and frequent progress monitoring (Horowitz, 2006). Curriculum-based measures (CBM; Deno, 1986; Deno, Marston, & Tindal, 1986) are a standardized form of curriculum-based assessments, created by Stan Deno at University of Minnesota, that provide brief, repeated skill assessments that are "flexible and reliable in showing where progress is and isn't happening, allowing teachers to intervene more effectively to make sure every child is getting what he or she needs to succeed" (University of Minnesota ResearchWorks, 2005).

In summary, both formative and summative assessments are important in analyzing the effectiveness of instructional interventions for young children. Current literature demonstrates promising implications for the use of progress monitoring to inform teaching practice and program evaluation. The ability to obtain valid and reliable evidence of student progress, or lack of progress, through quick and simple measures can be supportive of data-driven instructional decision-making, one of the hallmarks of good teaching. Formative and summative data can also be used to evaluate potential moderating effects of classroom peers and teachers on individual student achievement.

Conceptual Framework for Current Literature Review

Recent research has emerged that moves beyond simply relating early childhood programmatic features to classroom quality. This new body of research looks more closely at the relationship among student, teacher and classroom characteristics to student outcomes. This research is founded in ecological systems theory, an area of developmental psychology (Bronfenbrenner, 1974; Bronfenbrenner & Morris, 1998). Operating from an ecological perspective, educational theorists view the child and his or her environment as interconnected at the mesosystem level, with the child contributing to changes in the environment and the environment contributing to changes in the child. Thus, student, teacher and classroom characteristics combine as important factors that determine each individual student's academic outcomes.

Beginning with the child, it is necessary to consider the broader environment from which the child enters the classroom and engages in learning with the teacher and peers. Factors often considered include SES, language status, and race/ethnicity, those risk factors previously identified as highly predictive of academic failure (NCES, 2002; NRC, 1999). A child's home and family comprise his or her primary and constant environment at the microsystem level (Bronfenbrenner & Morris, 1998), and cannot be extricated from classroom learning outcomes. In addition, entering achievement and ability are inherent factors the child brings to the classroom and any learning opportunities he or she experiences there.

An ecological perspective also considers the inherent and external factors of a child's classroom peers as important features of the environment, which interact to impact the child's outcomes. Dotterer and her colleagues (Dotterer, Burchinal, Bryant,

Early, & Pianta, 2009) looked at the effect of classroom peer characteristics on student's outcomes, finding that children enrolled in more economically diverse pre-kindergarten programs achieved greater outcomes than children who attended programs consisting only of peers from low SES backgrounds. These findings coincide with earlier research showing that a child learning in a classroom consisting of greater numbers of poor children was more likely to encounter low quality learning opportunities than his or her peer in a classroom with children from higher SES backgrounds (Pianta, Howes, Burchinal, Bryant, Clifford, Early et al., 2005). In addition to students' SES status, an ecological perspective would consider the makeup of the classroom in terms of race/ethnicity, English language ability, and academic ability as intertwined to each individual student's learning outcomes.

A third main contributor to a child's present and future academic achievement is the classroom teacher. Operating from an ecological perspective, teachers are considered direct and indirect mediators of student learning (Justice & Sofka, 2010). Direct teacher input occurs via explicit interactions with children, while indirect inputs occur through the classroom environment and learning opportunities teachers set up for their students. Thus, the teacher is seen as primarily responsible for the interactions, environment and experiences of children in his or her classroom. Teacher-related factors such as his or her educational background and relevant training, teaching experience, and quality of interaction may have an effect on students' outcomes (Pianta et al., 2005).

Figure 1 illustrates a child's language and literacy outcomes mediated by his or her teacher and peers. This builds on the work of Justice and Sofka (2010), who posited that teachers' inputs act as mediators to children's early literacy understanding. This

ecological perspective applies to a host of outcomes for students, but in the context of this study the focus is on language and literacy outcomes, as researchers have demonstrated that children's early language and literacy ability are key predictors of their later academic outcomes (Lennon & Slesinski, 1999; Torgesen, 2002; Vellutino et al., 1998).

Overview of Research on Ecological Factors

Using research on the promise of early intervention, many states began to implement public early education programs. To ensure equity of opportunity for all students, a small number of states have implemented "universal pre-k" (UPK) programs that offer early education to all students meeting age requirements regardless of risk factors. Researchers have recently begun large-scale efforts to analyze the effect of state and federal pre-kindergarten programs on student cognitive, academic and social outcomes (Administration for Children and Families [ACF], 2002; Early, Barbarin, Bryant, Burchinal, Chang et al., 2005; Henry et al., 2001, 2003; Peisner-Feinberg & Maris, 2005a, 2005b; Zill & Resnick, 2006). The largest of these program evaluations are: (a) Head Start FACES study; (b) National Center for Early Development and Learning (NCEDL) Multi-State study; and (c) NCEDL SWEEP study.

Head Start Family and Child Experiences Survey (FACES). FACES is a longitudinal research effort to assess the effectiveness of Head Start, a federal program primarily for children living below the poverty line, across a nationally representative sample of Head Start participants, and included close to 5,000 three and four-year old children (ACF, 2006). Children whose parents applied to participate in Head Start were randomly assigned by lottery to a treatment group, in which students were enrolled in Head Start, and a control group, in which students were not enrolled in the program.

Classroom quality data were gathered using the revised Early Childhood Environment Rating Scale (ECERS-R; Harms, Clifford & Joyner, 1998). Teacher self-report data on the type and amount of academic instruction was reported (ACF, 2006). Child academic and social outcomes were measured, as well. Analyses showed that children enrolled in Head Start began at lower levels than a national peer sample across scales. And, while students in Head Start made statistically significant gains in vocabulary, early writing and math skills, and positive social behavior compared to comparison group children, children's skills remained well below national norms (ACF, 2006), suggesting that students enrolled in Head Start did not "close the achievement gap" with their more advantaged peers. The relationship between teacher quality and student outcomes was not assessed.

NCEDL Multi-State Study. NCEDL reviewed pre-kindergarten implementation, including Head Start, traditional public schools, and state-funded classes in community-based settings in six states – California, Georgia, Illinois, Kentucky, New York, and Ohio (Early, Barbarin, Bryant, Burchinal, Chang, Clifford et al., 2005). The goals of the study were to understand differences in pre-k programs and to ascertain if these differences led to differential child outcomes following pre-k and kindergarten. A total of 240 randomly selected sites participated in project over the course of the 2001-2002 school year, for a total of 960 students. The study found that 53% of enrolled students came from families earning less than 150% of the federal poverty guideline; minority students were more likely than white students to be enrolled in programs serving low-income students; and 42% of students had mothers with no more than a high school education.

Children made small but statistically significant gains on national assessments of receptive vocabulary, spoken language, and early math ability. Yet, students still performed below the national norm across all domains (Frank Porter Graham Institute [FPG], 2005). By disaggregating the data, researchers found that quality of emotional and instructional interactions between teachers and students were typically lower when the classroom served mostly poor children and teachers had less than a bachelor's degree in early childhood education (FPG, 2005).

State-Wide Early Education Programs (SWEEP) Study. Researchers conducted the Study of State-Wide Early Education Programs (SWEEP) during the 2003-2004 school year across five states – Massachusetts, New Jersey, Texas, Washington, and Wisconsin. These states were targeted in order to provide complementary data to those collected from the earlier *NCEDL* Multi-State Study. The states included in this study had diverse funding sources and delivery models (Early et al., 2005). This study included 1,840 students from 465 randomly selected public pre-kindergarten, Head Start, and state-funded pre-kindergarten classrooms in community-based settings across the five states. Findings were similar to those of the Multi-State Study across child outcome and teacher and classroom quality measures. Teacher and classroom quality characteristics, specifically quality of emotional and instructional interactions between teachers and children, were positively correlated with students' socioeconomic status. Finally, student gains across programs were statistically significant but not great enough to close the achievement gap with their more advantaged peers (Early et al., 2005).

Secondary analyses of large-scale studies

In addition to state and federally funded program evaluation studies, which demonstrate the effectiveness of an entire program on child outcomes compared to children not served by the program, more targeted peer reviewed analyses of these program evaluations have emerged that investigate within program differences that may influence variable effects for students with different characteristics who were all served by the programs evaluated (Curby, LoCasale-Crouch, Konold, Pianta, Howes, Burchinal et al., 2009; Early, Bryant, Pianta, Clifford, Burchinal, Ritchie et al., 2006; Early, Maxell, Burchinal, Ebanks, Henry et al., 2007; Gerde & Powell, 2009; Gallagher, & Lambert, 2006; Howes, Burchinal, Pianta, Bryant, Early, Clifford et al., 2008; Mashburn, Pianta, Hamre, Downer, Barbarin, Bryant et al., 2008; Missall, McConnell, & Cadigan, 2006). Rather than evaluating programmatic features that impact student achievement, these researchers analyzed outcomes at the teacher and student level to understand how teacher and student characteristics affect four-year olds' language or literacy achievement at the classroom and student level.

Summary. Program evaluation studies have analyzed whether or not students participating in the respective early childhood public education programs demonstrated more achievement than children not enrolled in the program and considered classroom level variables as well. Secondary analyses have evaluated the effect of teacher factors on four-year old children's outcomes using standardized, norm-referenced achievement tests. However, only one published study analyzes the relationship between teacher and student characteristics and students' change in performance on progress monitoring assessments (Missall, McConnell, & Cadigan, 2006), which is demonstrated as a more effective way to promote student achievement during a school year than outcome-only

analyses (Byrne, Fielding-Barnsley, & Ashley, 2000; Capizzi & Fuchs, 2005; Fuchs, Fuchs, Hamlett, & Stecker, 1991; VanDerHeyden, Witt, & Naquin, 2003; Vellutino, Scanlon, & Lyon, 2000). No studies look at how change in teacher quality over the school year might affect student language and literacy outcomes. Primary and secondary analyses for the reviewed studies were conducted in traditional public school and Head Start programs, whereas there are no published studies evaluating these relationships in nontraditional school choice programs such as charter schools.

Purpose and Possible Significance of Study

The purpose of this research was to address gaps in the current published literature related to peer and teacher quality effects on student language and literacy outcomes in early childhood classrooms. By analyzing data from both three and four-year old classrooms, including data for students with disabilities and second language learners, and using data from public charter schools, the study may provide important information about key factors influencing academic achievement for at-risk and typically developing students in urban charter school settings. In addition, this study aims to add to the previous research on point-in-time teacher quality effects by analyzing relationships between teachers' change in quality across a school year and students' language and literacy achievement. Measuring the difference in teachers' quality over time may be a more accurate reflection of children's classroom experiences during the school year than analysis of observed quality for a single point in the school year. The current focus on using preschool education as a method of intervention to enhance student achievement and to identify those students who need additional support underscores the relevance and potential importance of such a study. Furthermore, with

the increased focus on accountability for federally funded early education programs such as Head Start (Association for Children and Families [ACF], 2010), research on an existing federally supported program may help to inform policymakers and practitioners about early childhood practices that might support program goals and objectives.

Research Questions

This study addresses whether peer or teacher variables affect language and literacy formative or summative achievement, or both formative and summative achievement, or change in formative achievement for at-risk and typically developing three and four-year old children.

1. *Research Question (RQ) 1*: To what extent are selected classroom peer characteristics associated with students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?

2. *Research Question (RQ) 2*: What is the relationship between teacher characteristics and students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?
3. *Research Question (RQ) 3*: To what extent are selected classroom peer characteristics associated with students' language and literacy achievement and change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?
4. *Research Question (RQ)4*: What is the relationship between teacher characteristics and students' spring language and literacy achievement or fall-to-

spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?

These research questions were addressed by examining nationally normed and criterion language and literacy scores of 431 three and four year old students attending public charter schools in Washington, DC. Data for special education and ELL students were included in all analyses. Analyses from this research provided information about key peer and teacher characteristics influencing academic achievement for at-risk and typically developing students in urban charter school settings. This adds to the existing research on teacher and peer characteristics that moderate student achievement.

Definitions of Key Terms

Achievement. Students' demonstrated academic ability based on both norm- and criterion referenced tests of language and literacy development.

Achievement gap. The difference in academic performance between different ethnic and or economic groups (USDOE, 2009), specifically on standardized language or literacy measures.

Charter school. Charter schools are independent public schools of choice that receive per-pupil operational funding according to local and state formulas and regulations.

Classroom characteristics. Composite information about individual student characteristics for a classroom, along with descriptive information such as teacher to student ratio and implemented curriculum.

Criterion assessments. Criterion assessments are designed to assess student performance against mastery of a defined standard or objective. In the present study, criterion assessments refer to general outcome measures of student achievement rather than classroom-based assessments.

Language (Oral). The individual and connected vocabulary a child can comprehend (receptive; listening) and communicate (expressive; speaking) (Dunn & Dunn, 2007; Lonigan, Wagner, Torgesen & Rashotte, 2000; Williams, 2007).

Literacy. Pre-reading skills needed by children to become able readers - concepts about print (alphabet knowledge; letter-sound relationships; print and book knowledge), emergent writing, phonological and phonemic awareness (Lonigan et al., 2000; McBridge-Chang, 1998; Whitehurst, 2001; Whitehurst & Lonigain, 2001).

Norm-referenced assessments. Norm-referenced assessments are designed to assess students' performance in comparison to a range of demonstrated performance on the same assessment.

Student characteristics. Descriptive information about individual students including race/ethnicity, socioeconomic status (SES), language or literacy ability, English language background, and disability status.

Targeted Pre-Kindergarten (TPK). Pre-kindergarten program that is available to all four-year old children within a state who meet additional eligibility requirements, such as evidence of academic risk factors such as low socioeconomic status, minority classification, English as second language, or low achievement on entrance assessments.

Teacher characteristics. Descriptive information about teachers such as race/ethnicity, highest level of education, degree focus, specialized training in early childhood education, certification status, and years experience.

Teacher-child interaction. All verbal and non-verbal communication between adults and children in a classroom setting. Interactions may be related to academic learning, student behavior, or social in nature.

Teacher quality. Teachers' ability to provide high quality instruction as measured by observational tools.

Universal Pre-Kindergarten (UPK). Pre-kindergarten program that is available to all four-year old children within a state, regardless of income or other demographic data, through voluntary enrollment (Gormley, Gayer, Phillips & Dawson, 2005).

Chapter 2: Review of the Literature

According to Brofenbrenner and Morris (1998), children develop from an ecological perspective, where the child and his or her environment interact with one another and influence one another's development. Justice and Sofka (2010) applied this concept to language and literacy development, focusing on how the classroom environment helps to moderate student achievement. The following section presents a content and methodological review of the literature focused on identifying which teacher and classroom variables are associated with positive language and literacy outcomes for three and four-year old children. The goal of the review was to uncover common themes across studies, highlight potential gaps in the literature, and inform the dissertation study.

Reviewed studies are divided into three groups: (a) studies that evaluated the relationship between teacher background and student outcomes; (b) studies that evaluated the relationship between teacher and or classroom quality and student outcomes; and (c) studies that evaluated the relationship between peer characteristics and student outcomes. Finally, recommendations for future research are outlined based on findings from the content and methodological review of relevant literature.

Method

For this literature review, studies were retrieved using the University of Maryland's Research Port electronic resource library. The researcher searched for studies using the following databases: PsychINFO; Education Research Complete (EBSCO); ERIC; and Academic Search Premier. These keywords, in varying combinations, yielded relevant articles: *preschool quality; teacher qualifications; teacher knowledge; preschool; teacher variables; child variables; early intervention; early*

literacy; literacy; language; and *language development*. Upon finding potentially relevant articles, an ancestral search was conducted of those studies' references to yield additional studies for review.

Inclusion criteria. Articles were included in this review based on four criteria. First, only studies that were published in peer-reviewed journals were included. Second, retained studies were experimental, quasi-experimental, and non-experimental. Case studies and qualitative research studies were excluded. Combining the first and second criteria yielded 29 possible studies.

Third, the early intervention literature was narrowed to peer-reviewed experimental, quasi-experimental or correlational research that: (a) analyzed teacher and or classroom quality variables; (b) analyzed student outcomes in language or literacy; (c) examined preschool or pre-kindergarten; and (d) disaggregated outcomes related to students from commonly identified "at-risk" categories, such as children from low socioeconomic backgrounds, children who are members of racial minority groups, and children whose native language is not English (NCES, 2002; NRC, 1999; Snow Burns, & Griffin,1998).

Ten articles met all inclusion criteria and were retained for in-depth review. A summary of each study's content, organized alphabetically by author, is presented in Table 2.1. Descriptive information provided in the table includes (a) stated purpose of research, (b) participants, (c) design, (d) independent variables, (e) dependent variables, (e) results, and (f) limitations. To complete this table, the researcher used the Test et al. content matrix as a guide (Test, Fowler, Brewer & Wood, 2005). In regards to the four inclusion criteria, all 10 of the studies are non-experimental.

Content Analysis

All studies analyze data for pre-kindergarten classrooms, which enroll students between 4- and 5-years old depending on the study. Five of the ten studies performed secondary data analyses of data from the NCEDL Multi-State and or SWEEP studies (Burchinal, Howes, Pianta, Bryant, Early, Clifford et al., 2008; Curby et al., 2009; Early et al., 2006; Howes et al., 2008; Mashburn et al., 2008). A sixth study analyzed data from these two studies as well as five other studies of early education, including the Head Start FACES study (Early et al., 2007). One study analyzed data from Head Start programs across several states (Gallagher & Lambert, 2006). The final three studies used original data collection across a smaller group of classrooms and students (Brown, Molfese & Molfese, 2008; Gerde & Powell, 2009; Missall et al., 2006).

Combined, these studies analyzed data for a large number of classrooms, teachers and students across several states and programs. The results from these 10 studies are summarized for content related to: (a) the relationship between teacher background and student outcomes; (b) the relationship between teacher and or classroom quality and student outcomes; and (c) the relationship between classroom peers' characteristics and student outcomes. Tables 2.2 and 2.3 provide descriptive information on teacher, student and classroom variables analyzed across studies.

Teacher background. Teachers' educational background was considered in four of the ten studies (Brown et al., 2008; Early et al., 2006; Early, 2007; Gerde & Powell, 2009). However, among these four studies, data were collected and analyzed differently. Researchers reevaluated student outcomes related to teacher background characteristics, which included highest level of education, degree type, number of years' teaching, and

education credentials.

Early and colleagues (2006) sought to determine if teachers' education and or credentials predict children's academic gains. The researchers used a subset of data from the NCEDL Multi-State/SWEEP studies. Analyzing data for 237 classrooms and 878 pre-kindergarten children without identified disabilities, researchers operationalized teacher education according to years of education, highest degree, and bachelor's degree or no bachelor's degree. In addition, Early et al. (2006) examined the role of teachers' college major, teaching certification and or Child Development Associate (CDA) credentials on students' language and literacy outcomes. All teacher data were collected via questionnaire from participating teachers.

Students were assessed using a variety of measures. Students' receptive and expressive language ability were measured using the Peabody Picture Vocabulary Test 3rd Edition (PPVT-III; Dunn & Dunn, 1997) and the Oral and Written Language Scale (OWLS; Carrow-Wooldfolk, 1995) respectively. Phonological awareness ability was measured using the Woodcock-Johnson III rhyming subtest (WJ-III, McGrew, & Mather, 2001). Students' ability to identify upper and lowercase letters was measured using criterion-referenced NCEDL (2001) measures. Finally, students' ability to name colors was assessed using a measure from the Head Start FACES researchers (FACES, 1998).

Findings on the relationship of teachers' education, experience and credential to students' language and literacy outcomes were mixed. After controlling for state, fall scores, program type (e.g., school-based or community-based program), teacher-child ratio, school hours per week, and maternal education, there were two major findings related to language and literacy spring outcomes. Children whose teachers majored in

early childhood education or child development outperformed their peers on the color naming measure ($d = 0.23$, $p < 0.05$). When comparing teachers holding a CDA credential with teachers who held a high school diploma or associate's degree, holding a CDA credential was positively related to children's gains in letter identification, rhyming, and color naming. Education, training and credentials were not significantly related to gains in expressive or receptive language ability.

More recently, Early and colleagues analyzed data from seven of the eight large-scale early childhood education program evaluation studies to determine the effect of teachers' level and type of education on students' academic achievement across a variety of factors (Early et al., 2007). In all, findings represented data from 7,616 3- and 4- year -old children across 2,994 classrooms from 1995 to 2004. Teacher education was operationalized across four levels: (1) high school or general education diploma (GED); (2) associate's degree; (3) bachelor's degree; and (4) graduate degree. There were three categories used to describe teachers' academic major if they had completed at least a bachelor's degree: (a) early childhood education or child development; (b) other education major (e.g., elementary or special education); and (c) non-education major. In six of the seven program evaluations reviewed by Early et al. (2007), students' language ability was measured using the PPVT-III. Six evaluations also administered the WJ-III letter-word identification subtest.

Early and colleagues (2007) used multilevel modeling to analyze relationships at the classroom and student level. Controls at the classroom level included state, site, teacher-child ratio, class size, length of school day, teacher ethnicity, and proportion of students with low SES backgrounds. Child-level controls were for gender, ethnicity,

mother's education, SES background and initial assessment score (e.g., fall PPVT-III when analyzing spring PPVT-III).

Two of the seven program evaluations reviewed by Early and colleagues highlighted significant differences in students' pre-reading outcomes based on level of teacher degree (Early et al., 2007). That is, WJ-III letter-word identification scores were significantly higher for students in classrooms with teachers holding a bachelor's degree compared to those without a degree. Two evaluations also found a positive linear relationship between the four levels of education and WJ-III letter-word identification scores. Finally, one evaluation found a modest relationship between degree type and WJ-III letter-word identification scores, with student averages being higher in classrooms where teachers held an early childhood education or child development degree compared to those with degrees in non-education related fields. None of the program evaluation data demonstrated a significant relationship between teacher education level and students' language outcomes.

In a study of eight classrooms across three schools, Brown and colleagues (2008) analyzed the relationship between teachers' educational attainment and years of teaching experience and students' development of letter knowledge. Brown et al. examined teacher education as a dichotomous variable, comparing teachers with bachelor's degrees with teachers who had not obtained bachelor's degrees. The researchers also evaluated the relationship between years of teaching experience and student outcomes. Brown et al. collected and analyzed student data for 138 4-year old children from low-income families attending public pre-kindergarten programs.

The researchers (Brown et al., 2008) used the letter knowledge subscale of the

Wide Range Achievement Test (WRAT; Wilkinson, 1994) to assess children's ability to identify uppercase letters presented in random order. Findings were that both teacher education and experience moderately correlated with child literacy outcomes. After controlling for students' fall scores and teachers' literacy beliefs, teacher education was significantly related to students' letter identification outcomes. The relationship between teacher experience and students' letter identification was weaker but still significant.

Summary. Of the three studies that analyzed teacher background related to student language or literacy outcomes (Brown et al., 2008; Early et al., 2006; Early et al., 2007), findings were mixed. Teacher education, but not years of teaching experience, was statistically significant and modestly predictive of students' letter-word identification ability in two evaluations reviewed by Early et al. (2007). Brown et al. (2008) found that both teacher education and years' experience were moderately related to student outcomes in letter knowledge. For each of these studies, no relationships were found between teacher background and student outcomes in receptive or expressive language ability. In contrast, other program evaluation research has demonstrated improved language outcomes for students whose teachers have bachelor's degrees in early childhood education (Dotterer et al., 2009; Frank Porter Graham Institute [FPG], 2005). More research analyzing teacher background variables in relationship to student language and literacy outcomes may help tease apart the apparent discrepancies of research outcomes.

Teacher quality. Teacher quality was analyzed in relationship to student outcomes using five different measures across studies. Portions of the Classroom Assessment Scoring System (CLASS; LaParo, Pianta, Hamre, & Stuhlman, 2002) were used in four

studies. Portions of the Early Childhood Education Rating Scale – Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) were used in four studies. Researchers analyzed data for the Assessment Profile for Early Childhood Programs: Research Edition II (Assessment Profile: Abbott-Shim & Sibley, 1998) and the Ecobehavioral System for the Complex Assessment of Preschool Environments (ESCAPE; Greenwood, Carta, Kamps, Delquari, 1997) in one study each (Gallagher, & Lambert, 2006; Missall et al., 2006). Finally, a researcher-created tool measuring teacher verbal book-reading behavior was administered in one study (Gerde & Powell, 2009).

Classroom Assessment Scoring System (CLASS). The CLASS is an observational tool that uses a Likert rating from 1 to 7. Four of the 10 studies used multilevel modeling to analyze student language and literacy outcomes related to portions of the pre-kindergarten version of CLASS in the Emotional Support and Instructional Support domains (Curby et al. 2009; Early et al., 2006; Howes et al., 2008; Mashburn et al., 2008).

Emotional Support. On the version of the CLASS used in these studies, the Emotional Support domain measured teachers' support for students in the dimensions of positive climate, negative climate, teacher sensitivity, over control, and behavior management. Early and colleagues (2006) found no significant relationship between teachers' highest degree level and various measures of classroom quality. However, a small but significant and negative relationship was found between teacher licensure and quality of Emotional Support. Thus, teachers who were certified to teach 4-year old children within their state were rated as providing lower quality emotional support. No direct relationships between Emotional Support quality and students' language and

literacy outcomes were analyzed (Early et al., 2006).

In contrast, Howes and colleagues (2008) analyzed the relationship between quality of CLASS Emotional Support and students' academic gains. The researchers found that children made greater literacy gains in classrooms with higher ratings of Emotional Support. Using the NCEDL Multi-State/Sweep studies' data on nearly 3000 randomly selected pre-kindergarten children from 700 classrooms, the researchers analyzed the relationship between quality profiles of teacher-child interactions and pre-kindergarten children's achievement gains in receptive (PPVT-III) and expressive (OWLS) vocabulary and letter identification (NCEDL identifying letters test). Children showed larger gains on the NCEDL letter identification test when they experienced closer teacher–child relationships. Furthermore, controls were entered to demonstrate that gains were not related to characteristics of the child (i.e., state, gender, ethnicity, SES background, maternal education, and household size) or program (i.e., teacher degree, teacher to student ratio, length of school day, and setting). No relationship was found between Emotional Support and expressive or receptive vocabulary outcomes.

Mashburn and colleagues (2008) evaluated the relationship of quality of Emotional Support provided by teachers and students' language and literacy outcomes using the same NCEDL Multi-State/Sweep dataset as Howes et al. (2008). Child measures administered were the OWLS, PPVT-III, and WJ-III rhyming subtests. Data analysis involved multilevel modeling, with children nested within programs. At the child level, controls were entered for race/ethnicity, SES background, gender, and mother's level of education. At the program level, controls were entered for teachers' highest level of degree, field of study, class size, teacher to student ratio, utilization of comprehensive

curriculum, and whether or not the program provided meals, health or family services. An additional analysis was completed for the subgroup of Spanish-speaking ELLs (n = 283). For both the whole group and the subset of ELL students, there was no demonstrated relationship between quality of Emotional Support and children's development in language or literacy.

Curby et al. (2009) is the most recent study to analyze the relationship between quality of teachers' Emotional Support and children's receptive language outcomes. Also using multilevel analyses of the Multi-State and SWEEP data, researchers assessed the relationship between quality of Emotional Support and children's receptive vocabulary achievement gains in pre-kindergarten programs (Curby et al., 2009). After excluding children who did not have both fall and spring assessment scores on the measures of interest, a final sample of 2,028 students was obtained. Child level controls included in the analysis were age, days between fall and spring assessments, fall scores, gender, race, SES background, and maternal education. Classroom level controls included were dummy-coded CLASS profiles. Classrooms were configured into five different profiles: (1) high ratings across the three CLASS domains; (2) moderate Emotional and Organizational Support, high Instructional Support; (3) moderate Emotional Support; high Organizational Support, low Instructional Support; (4) moderate Emotional Support, low Organizational and Instructional Support; and (5) low ratings on all three CLASS domains (Curby et al., 2009). After controlling for the identified student predictors, researchers demonstrated that quality of Emotional Support was not significantly related to students' receptive language gains on the PPVT-III, the only language or literacy measure analyzed for this study.

Instructional Support. The four studies that analyzed the relationship between quality of teachers' Emotional Support and student outcomes in language and or literacy also analyzed student outcomes related to two dimensions of the Instructional Support domain, Concept Development and Quality of Feedback (Curby et al. 2009; Early et al., 2006; Howes et al., 2008; Mashburn et al., 2008). An additional study only evaluated the relationship between quality of Instructional Support and students' language and literacy outcomes (Burchinal et al., 2008).

Howes and colleagues' (2008) analysis found that children made slightly greater gains in receptive (d = .06) and expressive (d = .07) language measures when they experienced higher quality Instructional Support. Mashburn et al. (2008) used HLM on the same dataset with similar findings. Across language and literacy measures (i.e., PPVT, OWLS, WJ-III Rhyming, and NCEDL Letter Naming), students in classrooms with higher ratings of Instructional Support experienced statistically significant greater gains (Mashburn et al., 2008).

Also using multilevel analyses of the Multi-State and SWEEP data, Curby and colleagues assessed the relationship between quality of teacher-child interactions and children's receptive vocabulary achievement gains in pre-kindergarten programs (Curby et al., 2009). Researchers found that children in classrooms with the highest level of Instructional Support achieved higher language gains than children in classrooms with lower levels of Instructional Support. More specifically, children in classrooms with the highest level of support for Concept Development, one of the Instructional Support dimensions, had the highest gains in receptive language.

Burchinal et al. (2009) used multiple regression techniques on a subset of the data from the NCEDL Multi-State study to analyze the relationship between 240 randomly selected pre-kindergarten teachers' quality and nearly 700 students' language and literacy achievement at the end of the pre-kindergarten and kindergarten years. Students' expressive and receptive language were assessed using the OWLS and PPVT-III, respectively. Students were also administered the WJ-III letter-word identification subtest. Finally, phonological awareness ability was assessed using the WJ-III rhyming subtest and the CTOPP phonemic awareness subtest. OWLS and PPVT-III scores were compiled into a language variable, while WJ-III and CTOPP scores were compiled to create a literacy (reading) variable.

Researchers controlled for fall achievement, state, gender, ethnicity, ELL status, maternal education, and program hours per week. Effect sizes for gains on the language composite were modest but positive in favor of students from classrooms with higher quality of Instructional Support in pre-kindergarten (d = .08, p < .05) and kindergarten (d = .11, p < .01). Higher scores on the literacy composite were also predicted by students' placement in classrooms where teachers provided higher quality Instructional Support (d= .07, p < .05).

CLASS dimension and domain revisions. After these studies were conducted, the CLASS tool was revised to include ten dimensions across three domains, and the Emotional Support and Instructional Support domains now contain different dimensions. Specifically, behavior management is no longer considered a dimension of Emotional Support, and *Overcontrol* became *Regard for Student Perspectives*, with some rating item changes. In the domain of Instructional Support a new dimension was added, *Language*

Modeling. Thus, the CLASS authors have found that their original constructs were too broad in one domain and limiting in another. A third domain, Classroom Organization, was also added to the tool, which includes three dimensions: *Behavior Management*, *Productivity*, and *Instructional Learning Formats*. New research using the current version of the CLASS may demonstrate additional or different relationships than those found in previous studies, due to the construct changes and additions.

Early Childhood Education Rating Scale – Revised (ECERS-R). Two studies also used portions of the Early Childhood Education Rating Scale – Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) to predict student outcomes (Howes et al., 2008; Mashburn et al., 2009). ECERS-R is a measure of global classroom quality. Like the CLASS, quality is measured on a 7-point Likert scale, with 1 indicating "inadequate quality" while 7 indicates "excellent" quality.

Howes et al. (2008) found modest relationships between teachers' ratings on the ECERS-R Teaching and Interaction Scale and students' spring expressive (d = .06) and receptive (d = .07) language ability, after controlling for the same variables outlined in the review of CLASS outcomes. Finally, Mashburn et al. (2008) found higher scores on the ECERS-R positively associated with children's spring expressive language ability measured by the OWLS.

Other teacher quality measures. While the majority of studies used portions of the CLASS and or ECERS-R to determine teacher quality, three studies administered teacher quality measures that were only used once across the ten studies I selected for review (Gallagher & Lambert, 2006; Gerde & Powell; 2009; Missall et al., 2006).

Gallagher and Lambert (2006) used multilevel modeling to assess the relationship between concentration of children with special needs in general education classrooms and teacher quality for 600 Head Start children across 70 classrooms. Of the 600 students for whom data were included, 120 were identified as having one or more special needs based on a FACES parent report (1997). The authors assessed teacher quality using the Assessment Profile (Abbott-Shim & Sibley, 1998). This dichotomous observational checklist measures the presence or absence of evidence related to five scales: Learning Environment, Scheduling, Curriculum Methodology, Interacting, and Individualizing. Children were assessed using two subtests of the Metropolitan Early Childhood Assessment Program Pre-literacy Inventory (M-KIDS; Nurss, 1995), Print Concepts and Story Retelling.

Child–level controls were mother's education, SES background, gender, exposure to violence, maternal depression, child's age, and parental report of special needs (Gallagher & Lambert, 2006). Classroom-level controls were the classroom means for each of the child-level predictors. One key finding related to classroom quality. Classrooms with larger concentrations of children with special needs were rated higher in classroom quality than classrooms with smaller ratios of students with special needs. Children in classrooms where teachers had higher Assessment Profile ratings also scored higher on print concepts and story retelling measures.

Missall et al. (2006) compared the percent time spent in instructional situations measured by the Ecobehavioral System for the Complex Assessment of Preschool Environments (ESCAPE). ESCAPE is a direct observation system that measures teacher and student behaviors and classroom ecology. Unlike the other measures described here,

which analyze teacher quality within the entire classroom context, ESCAPE measures variables in relationship to an individual student. Thus, each child received his or her own ESCAPE rating.

Sixty-nine pre-kindergarten students' data were grouped into four categories: (a) 26 with mild disabilities; (b) 12 in Head Start with low SES; (c) 19 ELLs; and (d) 12 in family education programs who were not low SES, ELL or identified as having a disability (Missall et al., 2006). Researchers administered three subtests of the Individual Growth and Development Indicators (IGDIs; Priest et al., 2000a, 2000b) to evaluate student language and literacy development: picture naming, rhyming, and alliteration. Multilevel modeling was implemented to evaluate the relationship between ESCAPE ratings and students' outcomes on the IGDI language and literacy progress monitoring subtests. At the child level, ESCAPE ratings and IGDI measures were entered. At the group level, ESCAPE means and IGDI means were entered. No other control variables were entered into the model. Missall et al. (2006) found that ESCAPE ratings were significantly correlated with children's growth across measures.

Gerde and Powell (2009) analyzed whether an early childhood specialization contributed to variance in student outcomes, mediated by specific book-reading behaviors. Thus, Gerde and Powell (2009) sought to understand "the extent to which teachers' large group book-reading practices are a pathway between teachers' educational background and children's growth in receptive vocabulary" (p. 218). This study included 341 Head Start children from 60 classrooms within six centers in the Midwest. All students were identified as at risk for school failure due to their SES backgrounds, while 17% were ELL, an additional risk factor (NCES, 2002). Teachers' book-reading

behavior was captured using a researcher-created coding system designed to categorize teacher talk during whole group storybook reading into book or behavior related utterances. Students were assessed using the PPVT-III.

Using multilevel modeling, Gerde and Powell (2009) demonstrated that students in classrooms where teachers had more book-related utterances had higher growth levels on their PPVT-III scores (Gerde & Powell, 2009). Specifically, children gained 0.5 standard score points on the PPVT for each additional book-related utterance from their teacher. Findings also indicated that students who entered with lower vocabulary ability made significantly greater gains on the PPVT than their peers who began the school year with greater receptive language ability. Thus, children who began with lower scores decreased the gap in receptive vocabulary ability by the end of the year, although students who began the year with higher receptive vocabulary still scored higher at outcome.

Summary. Only one of the four studies analyzing the relationship between quality of Emotional Support measured by the CLASS and student language and literacy outcomes found any significant relationship, likely due to the limited variability in ratings across classrooms for this domain. Howes et al. (2008) found quality of Emotional Support positively related to gains in students' ability to identify letters. In contrast, students in classrooms where teachers rated higher on the CLASS Instructional Support domain demonstrated greater gains in language and literacy domains in each of the studies evaluating this relationship (Burchinal et al., 2008; Curby et al. 2009; Howes et al., 2008; Mashburn et al., 2008).

The ECERS-R observation tool was used in four studies. Only two of those studies directly analyzed the relationship between teachers' ratings on the ECERS-R and student

language or literacy outcomes. Howes et al. (2008) found significant positive relationship between ECERS-R quality and gains in expressive and receptive language ability. Mashburn and colleagues (2008) found a significant relationship for expressive language only.

The three studies using other teacher quality measures all found positive relationships between teacher quality and student outcomes. Outcomes demonstrated that children in classrooms where teachers had higher literacy-specific instructional quality also scored higher on: (a) print concepts and story retelling measures (Gallagher &Lambert, 2006); (b) expressive vocabulary and phonological awareness ability (Missall et al., 2006); and (c) receptive language ability (Gerde & Powell, 2009). Combined, the studies of teacher quality intimate that teachers' emotional interactions with students may contribute to increases in basic skills like letter identification, while the quality of teachers' instructional interactions may support students more general ability in expressive and recessive language. More research is needed that analyzes the relationships between general teacher-child interaction quality and student language and literacy outcomes and academic teacher-child interaction quality and student language and literacy outcomes, preferably with the same sample of children and teachers, to allow for more inference into relative strength of relationships.

Likewise, implementation of a high quality, scripted curriculum with fidelity may increase teachers' observed instructional quality. So, a requirement to implement a scripted curriculum in some programs may have masked potential differences in teacher-child interactions among teachers with varying amounts of education and teaching experience. None of the studies controlled for the possible effect of program

requirements such as standard curriculum implementation on students' outcomes.

Finally, all of the literature reviewed analyzed teacher quality at one point in the school year. None of the studies controlled for potential effects of professional development on teachers' observed quality, nor how changes in quality might impact students' language and literacy achievement. Previous research demonstrates that ongoing professional development can change teachers' performance and quality (Joyce & Showers, 1983; Neuman, 2009; Neuman & Cunningham, 2009) and impact students' language and literacy achievement (Powell, 2010). Additional studies analyzing relationships between change in teacher quality for teachers experiencing ongoing professional development and students' language and literacy achievement would add to this limited body of research.

Student and peer characteristics. There are a variety of classroom and student characteristics that might impact students' language and literacy development. Some of these variables include years of program exposure; baseline performance on measures of interest; ratio of students to teacher; race or ethnic makeup of class; number of ELL and/or special education students per class; children's SES background; and amount of time spent on instruction. Six of the ten studies controlled for the effect of gender, race or ethnicity, and SES at the child and classroom level on students' outcomes at the child and classroom level (Curby et al., 2009; Early et al., 2007; Gallagher & Lambert, 2006; Howes et al., 2008; Mashburn et al., 2008; and Missall et al., 2006).

Of the 10 studies reviewed here, only one analyzed the relationship between proportion of special education students in a classroom and students' academic gains (Gallagher & Lambert, 2006). Using data from 600 children in preschool and pre-

kindergarten classrooms across 70 Head Start classrooms, including 120 children identified as having special needs, Gallagher and Lambert's analysis demonstrated that the presence or absence of special education students in classrooms did not significantly impact children's outcomes on the M-KIDS Print Concepts or Story Retelling subtests. However, findings were different when the authors divided classrooms into two categories: classrooms having 1 to 20% of children with special needs and classrooms having more than 20% of children with special needs. Children in classrooms with greater than 20% children with special needs scored lower on a measure of print concepts than their peers in classrooms with fewer than 20% children with special needs.

Missall et al. (2006) used multilevel modeling to evaluate the variance in language and literacy skill growth using progress monitoring tools for different groups of students based on teacher, student and classroom characteristics. Unlike other studies reviewed here that used multilevel modeling, students were not nested in classrooms but in risk categories. So, students in the same classroom could be placed in different groups, and data for a single student might also be analyzed in more than one group (e.g., Head Start and ELL or Head Start and special education). There is no indication that the researchers disentangled potential effects of inclusion in more than one group for analysis. Correlations between subtest and ESCAPE category differed for the four groups: (1) typically developing students in family education (ECFE) programs; (2) ELL (3) special education (speech-language disabilities); and (4) low SES background (Head Start) students.

Children in the family education programs scored highest and had the largest slopes for all three IGDI subtests; rate of growth was similar for ELL and SLD children

in Picture Naming and Rhyming. Students from low SES backgrounds demonstrated the least growth and lowest outcomes across measures. Thus, students with no identified risk factors significantly outperformed their comparison groups in expressive vocabulary and phonological awareness ability. Also important, poverty as a risk factor was more associated with limited growth and lower outcomes than either ELL or special education status.

After Early et al. (2007) analyzed data from seven large scale pre-kindergarten program evaluation studies, only children from the NCEDL Multi-State/SWEEP evaluation demonstrated significant variance in prereading outcomes based on any of the student level controls. Early et al. (2007) found that for teachers with specialized training in early childhood education or child development, prereading outcomes were similar despite students' SES background. However, in classrooms where teachers held a degree other than education, poor students actually made more growth than peers from higher socioeconomic backgrounds.

Using data from NCEDL Multi-State/SWEEP, Howes et al. (2008) found that "children from a variety of different family structures, economic, racial, and ethnic groups benefited equally (or not) from pre-K participation" (p. 45). In contrast, Mashburn and colleagues (2008) also used data from the NCEDL Multi-State/Sweep studies with different results. Their analysis demonstrated that ethnic minority students scored lower and had less growth than their white peers across academic measures, while poverty was significantly and negatively related to all language and literacy outcomes except letter naming. Mashburn et al. (2008) found gender significantly related to student outcomes for two literacy measures, rhyming and letter naming, with girls

making greater gains between baseline and outcome assessments than their male peers, and outperforming boys at outcome.

Curby et al. (2009) assessed the relationship between pre-kindergarten quality profiles of teacher-child interactions and children's achievement gains using existing data from the NCEDL Multi-State/Sweep studies. In contrast to Mashburn et al. (2008), Curby and colleagues found no significant differences related to language or literacy outcome measures for gender. However, non-White students and those with lower SES generally scored lower than their white and more affluent peers across outcome measures.

Summary. Across studies, a variety of student, teacher and classroom characteristics were analyzed (see Tables 2.2 and 2.3). Six studies controlled for the effect of gender, race or ethnicity, and SES at the child and classroom level on students' outcomes at the child and classroom level (Curby et al., 2009; Early et al., 2007; Gallagher & Lambert, 2006; Howes et al., 2008; Mashburn et al., 2008; and Missall et al., 2006). One study, Gallagher and Lambert (2006), analyzed the relationship of proportion of special education students in a classroom and student language and literacy outcomes. The authors found that greater percentage of students in a classroom was associated with lower print concepts (Gallagher & Lambert, 2006). None of the studies analyzed the relationship between the number or percentage of ELL students in a classroom, or number of years of program exposure, and students' academic gains.

Poverty was found to be a significant factor in 4 of 10 studies. Missall et al. (2006) found poverty associated with less growth and lower outcomes in expressive language and phonological awareness ability than similar age peers identified as ELL, special education with speech language disability, or typically developing peers with no

identified risk factors. Early et al. (2007) found SES background to be significant in only one of the seven program evaluations reviewed, on only the WJ-III letter-word identification subtest, and only when teachers had a college major other than education. Interestingly, this relationship favored students from lower SES backgrounds (Early et al., 2007). Both Mashburn et al. (2008) and Curby et al. (2009) found that non-white students and those with lower SES generally scored lower than their white and more affluent peers across outcome measures. Additional research could analyze potential impact of key classroom peer characteristics demonstrated to affect student academic achievement: entering student achievement, as well as ELL and special education status.

Content summary. Only one of the ten studies (Brown et al., 2008) analyzed number of years' teaching experience in relationship to student outcomes, and found years' experience related to gains in students' letter knowledge. Findings were mixed related to the effect of teachers' education on classroom quality and student outcomes. However, teacher education was most consistently related to student outcomes in basic literacy skills (i.e., letter identification).

Across studies, student and classroom characteristics were not analyzed at the same level consistently. Specifically, although the majority of studies analyzed important child demographic variables at the classroom level (Burchinal et al.; Curby et al., 2009; Early et al., 2007; Gallagher & Lambert, 2006; Howes et al., 2008; Mashburn et al., 2008; Missall et al., 2006), three studies analyzed mean child characteristics across all participating classrooms rather than analyzing possible variance within classrooms using nesting (Brown et al., 2008; Early et al., 2006; and Gerde & Powell, 2009). Race/ethnicity, gender and SES were commonly analyzed variables, and each was found

to significantly impact outcomes in at least one of the studies. None of the three variables were found to be significant across studies.

In addition, average student achievement within and across classrooms was rarely considered. In fact, six of the ten studies (Burchinal et al., 2009; Curby et al., 2009; Early et al., 2006; Early et al., 2007; Howes et al., 2008; Mashburn et al., 2008) used data from the NCEDL Multi-State/SWEEP studies, which specifically excluded children with identified disabilities from the sample. ELL and special education students can be expected to perform more poorly than their English-only and typically developing peers. The relationships between ELL and special education status and individual student outcomes were only analyzed in two studies (Gallagher & Lambert, 2006; Missall et al., 2006). Although ratio of special education students per class was demonstrated as a significant factor in one study (Gallagher & Lambert, 2006), possible effects of number or percentage of ELL students at the classroom level were not analyzed in any study. No studies analyzed the potential effect of prior years' program experience on students' outcomes, which might be related to prior achievement.

Related to student language and literacy outcomes, a variety of measures were used. Student ability was assessed in expressive and receptive language, print concepts, and letter identification. No single study analyzed all of these measures, however, and relationships between teacher, student or classroom variables and student outcomes across these measures are unclear. Across studies using the CLASS, relationships were found between quality of teacher-child interactions and students' language and literacy outcomes. As the CLASS tool has been revised to include new constructs, as well as grouping dimensions into domains differently than was done at the time of data collection

for the reviewed studies, more research using the new protocols is needed to determine if these relationships remain.

The only study to analyze relationship between language and literacy-specific teacher-child interactions and students' language and literacy development used a single receptive vocabulary instrument and a researcher-created measure of teacher-child interaction quality that had not been validated on other populations (Gerde & Powell, 2006). Research that uses a validated tool to analyze the relationship between language and literacy specific teacher-child interactions and students' language and literacy development on multiple measures is unique and potentially important.

Furthermore, different studies identified positive relationships between quality of teacher-child interactions and students' achievement in receptive and expressive language, phonological awareness, and basic skills such as letter ID and print knowledge. No single study using the same sample has analyzed all of these measures for both general and language and literacy specific teacher-child interaction quality. Finally, when HLM was used to evaluate students nested in classrooms, only baseline and outcome language and literacy measures were used, and teachers' quality was analyzed at one point in the school year (Curby et al., 2009; Early et al., 2007; Gallagher & Lambert, 2006; Howes et al., 2008; Mashburn et al., 2008). Research analyzing relationships between the change in quality of teacher-child interactions and students' achievement, or change in achievement, on both norm- and criterion-referenced language and literacy measures is needed.

Methodological Analysis

Following is a methodology review of the 10 studies previously compared for content, to analyze strengths and limitations of the literature. Studies are analyzed according to how the researchers addressed four validity domains: (a) internal validity; (b) construct validity; (c) statistical conclusion validity; and (d) external validity. Each domain has several possible threats researchers must address in order to maximize a study's significance and generalizability, though different threats are of greater or lesser plausibility depending on type of study (Shadish, Cook, & Campbell, 2002). All of these studies are quasi-experimental with control groups and implemented pretests and posttests. Half of the studies included in this review reanalyzed data from randomized control trials studies (Burchinal et al., 2009; Curby et al., 2009; Early et al., 2006; Early et al., 2007; Howes et al., 2008; Mashburn et al., 2008), but the randomization was lost in the process.

After reviewing the selected studies, 8 of the 37 validity threats identified by Shadish et al. (2002) seemed potentially significant and plausible to this corpus. Operational definitions for the validity threats I selected to analyze are presented in Table 5. In this section, methodology is analyzed for three key areas described across the 10 studies: (a) teacher characteristics; (b) teacher quality; and (c) classroom and peer characteristics.

Internal validity. Internal validity refers to the ability of a researcher to make inferences about the cause of effect in a study by ruling out other possible causes (Shadish et al., 2002). To assess if the 10 selected studies adequately addressed the domain of internal validity, the researcher focused on three of the nine internal validity

threats identified by Shadish et al. (2002) as most relevant to this body of literature: (a) selection bias; (b) ambiguous temporal precedence; and (c) testing.

Selection bias. Failure to account for the possible effect of curriculum may contribute to selection bias, because the majority of studies analyzed data from large public programs that often mandate the use of specific curricula. Depending on fidelity of implementation, selecting programs that use a standard curriculum across classrooms may mask the potential effects of teacher education and years' experience, unnaturally making teachers' inputs more uniform across observations.

Other threats to selection bias are present due to studies' failure to account for the possible relationship between classroom concentration of ELL and/or special education students and student outcomes. Only one of the ten studies adequately controls for concentration of special education students within classrooms (Gallagher & Lambert, 2006). Finally, 6 of ten studies used data from the NCEDL Multi-State/SWEEP evaluations. Students with identified special needs were excluded from participation in both studies (Early et al., 2005, 2006, 2007).

Selection bias (maturation). Another threat to these studies' internal validity is maturation bias, as studies did not adequately report information regarding time of initial assessments. Specifically, ambiguous temporal precedence occurs when uncertainty about the order of introduction of dependent variables (e.g., language and literacy outcomes) and independent variables (e.g., classroom learning experiences) makes it difficult to ascertain the direction of the relationship. In these studies, this threat to validity was present because students may have had exposure to material measured by the assessments due to lapses between the beginning of the school year and baseline

assessments. This makes it unclear if students' baseline performance on the assessments was influenced by teacher instruction. In addition, controls were not consistently addressed for time between baseline and outcome assessments for those studies that administered measures at two time points.

As one example, fall and spring assessment administration averaged nearly 151 days apart in Early et al. (2006), but ranged from 91 to 244 days apart. The standardized measures administered were not standardized for administration in three-month windows, and administration so close in time may have contributed to testing bias. Students were only assessed at one time point in the only study to analyze the relationship of proportion of special education students in a classroom to language and literacy ability (Gallagher & Lambert, 2006). So, adequate controls for prior ability were not included.

Maturation is also a threat because none of the studies provide specific information regarding the timeframe for administration of the teacher and classroom quality assessments. In three studies (Curby et al., 2009; Early et al., 2006; Missall et al., 2006), classroom quality data were collected at only one time point, so it is unclear at what level of quality teachers' operated toward the end of the year, or how teachers' growth or outcome levels in quality predicted students' academic outcomes. Howes, et al (2008) and Mashburn and colleagues (2008) analyzed data where the CLASS was administered in both the fall and spring, but data are not provided on length of time between administrations.

Across all 10 studies, assessments administered by classroom teachers outside of study measures were not reported. Teacher administered assessments may have contributed to testing bias, where repeated exposure to an assessment measure may

impact subsequent scores, making it difficult to ascertain the effect of an independent variable. This information seems especially important for the criterion-referenced assessments such as letter and color identification, measures commonly explicitly taught and assessed repeatedly in structured pre-kindergarten programs. Seven of the ten analyzed studies used such assessments (Brown et al., 2008; Burchinal et al., 2008; Early et al., 2006; Early et al., 2007; Howes et al., 2008; Mashburn et al., 2008; Missall et al., 2006).

Construct validity. Construct validity refers to the researchers' ability to adequately define and assess the unit of interest, and may pertain to a treatment, sample, observation, or outcome measure (Shadish et al, 2002). If a study's constructs are ill defined, it is difficult to make inferences about how well any individual components of that study apply to the intended broader category or concept. Two of the 14 threats to construct validity identified in Shadish et al. (2002) seemed most relevant to this corpus: (a) mono-operation bias and (b) confounding of constructs.

Mono-operation bias. Mono-operation bias refers to a study's construct definition being inappropriately limited by the use of a single assessment measure, or the limited scope of multiple assessment measures (Shadish et al., 2002). Teacher and or classroom quality was most frequently assessed using a single instrument, with five of the ten selected studies using results from one observational assessment tool as a proxy for teacher or classroom quality (Curby et al., 2009; Early et al., 2007; Gallagher & Lambert, 2006; Gerde & Powell, 2009; Missall et al., 2006). With the exception of one study (Gallagher & Lamber, 2006), the measures used in this corpus demonstrate sound

psychometric properties having been validated by use in multiple classrooms across the United States.

Students across studies were assessed using only single measures of receptive or expressive language, letter knowledge and or phonological awareness. The studies did employ language and literacy measures that demonstrate validity and reliability. Reliance on a single measure to evaluate a construct contributes to mono operations bias. While several studies did implement multiple measures of language and literacy, no studies assessed the same group of students on different measures of a single construct (e.g., two or more measures designed to assess students' receptive language ability).

Confounding of constructs. Operations usually involve more than one construct, and failure to describe all constructs may result in invalid inferences. Four studies used two observational assessments (Early et al., 2006; Early et al., 2007; Howes et al., 2008; Mashburn et al., 2008). Researchers used portions of the ECERS-R to predict student outcomes; however, it is unclear which items were included (ECERS-R; Harms, Clifford, & Cryer, 1998). Use of this tool as a measure of teacher quality might violate construct validity. First, this observational tool measures classroom quality more globally, with many items not directly related to teacher inputs. In addition, the studies used different items from the ECERS-R, and it is not clear which items were discarded or maintained across measures. Therefore, it is unclear if the items evaluated relate to teacher quality or other constructs such as environment or program quality. This uncertainty also contributes to possible mono operation bias, as the selected items might not adequately measure the intended construct or measure unintended constructs.

Statistical conclusion validity. Nine threats to statistical conclusion validity were

identified by Shadish et al. (2002). Statistical conclusion validity refers to whether or not the independent and dependent variables covary and how strong the relationship of covariance truly is. Of the nine threats, the studies in this review seemed most vulnerable to two: (a) restriction of range; and (b) unreliability of treatment implementation.

Restriction of range. For those studies using the CLASS Instructional Support domain as a measure of teacher quality, a major threat to statistical conclusion validity is restriction of range. Restriction of range occurs because reduced range on a variable typically serves to weaken the relationship between it and another variable. CLASS uses a seven-point metric to assess quality in each domain. In all of the studies, however, average teacher performance on the Instructional Support domain remained in the low range. Burchinal et al. (2009) reported spring Instructional Climate score $M = 2.55$ ($SD = 1.10$). Howes et al. (2008) reported Instructional Climate data for the NCEDL Multi-State study (M = 2.20 [SD = .79]) and SWEEP study (M = 2.00 [SD = .87]). These Instructional Climate averages apply to all studies re-analyzing the relationship between teachers' rating of Instructional Climate and student academic outcomes using data from the NCEDL Multi-State/SWEEP studies (Curby et al., 2009; Mashburn et al., 2008). In fact, only 13% of teachers in the NCEDL Multi-State/SWEEP classrooms scored at or above 3 on the CLASS Instructional Support domain (Hamre, Goffin, & Kraft-Sayer, 2009). Because of the limited range of CLASS scores for the sample of teachers, the relationship between teacher quality and student outcomes might be underestimated.

Unreliability of treatment implementation. Unreliability of treatment implementation occurs when a treatment is not implemented in a standardized manner, or is only implemented in part for some participants (Shadish et al., 2002). When this

occurs, there is a risk that the strength or nature of relationship between the independent and dependent variables might be incorrectly estimated. Pertaining to this group of studies, *treatment* refers to participation in an evaluated preschool or pre-kindergarten program or classroom. Across all studies, participants received treatment in a variety of settings, and studies did not address the various program characteristics that might impact student outcomes, such as curriculum and access to materials. The lack of reliability in treatment implementation also threatens the studies' external validity (see *generalizability of treatment* below).

External validity. The final type of validity researchers must ensure in order to appropriately assess significance of their research efforts is external validity. External validity refers to the ability of the research outcomes to generalize to a broader or varied participant group, setting, treatment, or outcome measure (Shadish et al., 2002). Shadish and colleagues (2002) identified five threats to external validity. The three threats most relevant to these studies are: (a) generalizability of participants (*interaction of the relationship over units in);* (b) generalizability of treatment (*interaction of the relationship over treatment variations*); and (c) context-dependent mediation.

Generalizability of participants. One threat to external validity is generalizability of participants, when a relationship determined with one sample might not transfer to a different sample. Six of the ten studies specifically excluded students with identified disabilities from study participation (Burchinal et al., 2008; Curby et al., 2009; Early et al., 2006; Early et al., 2007; Howes et al., 2008; Mashburn et al., 2008). As a matter of law, all students are afforded the opportunity to receive a free and appropriate public education in the least restrictive environment (IDEA, 2004). Therefore, findings relating

teacher and classroom characteristics to outcomes for typically developing students, even those at risk for later academic problems, might not generalize to special education students. These findings also might not carry over into classrooms that include students with special needs.

In addition, nine of ten studies used samples from states or regions with lower percentages of African American and Hispanic students than typically found in poor, urban environments (NCES, 2008), and racial and ethnic makeup of study participants were not representative of these urban populations (see Table 4). These studies might provide limited application for urban environments with a majority of African American, Hispanic and other ethnic minority students.

Generalizability of treatment. External validity is also threatened by restricted generalizability of treatment. As noted previously, treatment refers to participation in the analyzed preschool and pre-kindergarten programs. A threat to generalizability of treatment occurs when a relationship found with one treatment might not be maintained with treatment variations, combinations of treatments, or partial treatments (Shadish et al., 2002). In this case, none of the ten studies sufficiently described the pre-kindergarten programs from which each sample was derived. Information regarding curriculum implementation, structure of classrooms or daily schedules, and other pertinent information was not provided. It is not clear, then, for which programs the relationships between teacher, classroom and student characteristics and student language and literacy outcomes found in these studies would generalize.

Context-dependent mediation. A threat to external validity that is increased due to limited generalizability of participants and treatment is the threat of context-dependent

mediation. This occurs when a correlational mediator in one context or setting might not predict outcomes in another context (Shadish et al., 2002). Across studies, because of limited variability in participants and or limited explanation of treatment, it is unclear if the relationship of teacher characteristics, such as education and years' experience or teacher quality to student outcomes, would persist in different contexts or with different student populations.

Methodology summary. Selected studies were analyzed according to how the researchers controlled for four validity domains: (a) internal validity; (b) construct validity; (c) statistical conclusion validity; and (d) external validity. Key threats to validity were identified related to teacher, student, classroom, and measurement variables across studies. Future studies that take these methodological factors into account in their design may lead to content findings different or more nuanced than those demonstrated in the ten studies reviewed here.

Summary of Content and Methodology Reviews

The preceding literature review provided several possibilities for what data analysis in the proposed study might demonstrate. Based on the review of literature, the hypotheses guiding this study are that:

- students' SES background will be less predictive of academic outcomes than teacher and classroom peer characteristics (Early et al., 2007);
- students' entering academic ability will be highly predictive of students' performance, with higher rates of growth for ELL, special education, and students in their first program year (Gallagher & Lambert; 2006; Missall et al., 2006);

- teacher characteristics most predictive of student outcomes will be quality of teacher-child interactions (Burchinal et al., 2008; Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008);
- teacher-child interaction measures that are targeted to language and literacy interactions will be more predictive of students' language and literacy achievement than general teacher-child interaction measures (Burchinal et al., 2008; Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008);
- teachers' degree type will show little relationship to student outcomes (Brown et al, 2008; Early et al., 2006; Early et al., 2007); and
- classroom characteristics most predictive of student outcomes will be peer characteristics, specifically baseline academic and cognitive ability (Gallagher & Lambert, 2006; Missall et al., 2006).

There are myriad studies supporting the importance of early intervention for at-risk students' later academic achievement. Less studied is how student, classroom and teacher characteristics individually and combined impact language and literacy outcomes for various at-risk student groups, specifically three and four-year old ELL and special education students. No known studies analyze pre-kindergarten student outcomes in public charter programs, and none analyze achievement or change in achievement for ELL or special education students from an ecological perspective (i.e., students nested within classrooms). No studies analyze how change in teacher quality relates to students' achievement or change in achievement on norm- or criterion-referenced language and literacy assessments. Future research should address these key gaps in the literature.

Discussion and Rationale for Study

The purpose of this research was to analyze the effectiveness of early language and literacy intervention efforts within the context of publicly funded charter preschool and pre-kindergarten classrooms serving students identified as ELL, special education and low SES. The current focus on using early childhood education as a method of intervention to enhance student achievement, and to identify those students who need additional support, and the continued understanding of the importance of early language and literacy ability to later academic achievement underscore the relevance and potential importance of such a study.

Support for UPK continues to increase, and policy makers are poised to begin enacting legislation expanding early education services for all students. It is therefore prudent that data be synthesized and reexamined to highlight how those dollars might best be spent to impact language and literacy outcomes. New studies that control for variables such as curriculum, the amount and type of language and literacy related professional development provided, and hours of daily instruction will add more light to the core teacher and classroom variables most related to positive student outcomes (Early et al., 2007; Howes et al, 2008). Furthermore, analyzing these data against students assessed using the same measures at the same time points will also add to the research (Gerde & Powell, 2009; Howes et al., 2008).

Research using the newly revised CLASS tool, which includes additional constructs not measured in the previous studies, will also contribute to an understanding of the relationship between teacher quality and student outcomes. In addition, the ability to look at if change in ratings for CLASS measures impacts students' language and

literacy development further advances the literature on this widely administered tool. Pairing the CLASS with a related and validated teacher quality measure, the ELLCO, will help to address mono operations bias. Studies using these assessments should address controls for administration across classrooms, including training, assessment windows, and analyzed items. All of the studies analyzed quality of teacher-child interactions at one point in the school year rather than analyzing change in teachers' observed quality from different points in the year (e.g., fall and spring observations), "which may limit the degree of associations between PK experiences and child outcomes" (Burchinal et al., 2008, p. 152).

Moreover, the research available mainly analyzes data for four-year old students, and does not account for differences in years of prior program participation. The body of research would benefit from the ability to analyze variability in growth for students younger than four in their first year of a program, four-year old students in their first year of a program, and four-year old students in their second year of a program (Curby et al., 2009). Only one study (Missall et al., 2007) used progress-monitoring assessments to analyze the relationship between teacher quality and changes in students' language and literacy achievement over time. Missall and colleagues analyzed progress monitoring for a limited sample of 69 students, but only used one assessment and nested students in conditions (i.e., special education, ELL) rather than classrooms. This study builds on their work by analyzing data for 560 students, using additional assessments to address mono-operations bias, and nesting students in classrooms as well as accounting for student characteristics. Finally, all of the available research analyzes student data from publicly funded traditional pre-kindergarten or Head Start classrooms serving four-year

old children. There is no research that includes three and four-year old students, and classroom peer and teacher-child interaction variables related to student outcomes in the relatively new but fast growing publicly funded charter school movement.

Chapter 3: Methodology

The preceding literature review provided information about student, teacher and peer characteristics predictive of preschool students' language and literacy achievement. Gaps in the literature were also identified. The current study aimed to address specific gaps through use of a multilevel design.

Research Design

The research was designed to determine if teacher and classroom peer characteristics predict student achievement, or change in achievement, after controlling for individual student demographic and prior achievement variables. This dissertation research was modeled after Lee and colleagues' two-level hierarchical linear modeling (HLM) design (Lee, Burkam, Ready, Honigman & Meisels, 2006), with children nested within classrooms rather than schools (see Figure 2). Level 1 includes within-classroom statistical controls for children's social and academic backgrounds and entering academic achievement. Level 2 includes classroom-level controls for teacher background, teacher behaviors, and academic composition. HLM allows the researcher to analyze the extent to which classroom-level variables, such as peer and teacher characteristics, explain variance in achievement not explained by student-level variables, such as demographics and prior achievement (Hoffmann, 2004). These relationships are explored through four research questions that focus on peer and teacher effects on global summative assessments and curriculum-based formative assessments.

Research questions and hypotheses. The research questions were paired based on the assessment type to be analyzed.

Research questions using norm-referenced, global language and literacy assessments. Questions one and two explore the relationships between peer and teacher variables and students' outcome achievement on standardized, nationally norm-referenced language and literacy assessments. These assessments are global measures of language and literacy normed for use with young children through adults, and are not designed to be curriculum-sensitive.

1. *Research Question (RQ) 1*: To what extent are selected classroom peer characteristics associated with students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?

   H_{01}: Classroom peer characteristics do not affect student language or literacy outcomes $\Delta r^2 = 0$

   H_{A1}: Classroom peer characteristics affect student language or literacy outcomes $\Delta r^2 \neq 0$

2. *Research Question (RQ) 2*: What is the relationship between teacher characteristics and students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?

   H_{02}: Teacher characteristics do not affect student language or literacy outcomes $\Delta r^2 = 0$

   H_{A2}: Teacher characteristics affect student language or literacy outcomes $\Delta r^2 \neq 0$

Research questions using criterion-referenced, curriculum-based language and literacy assessments. Questions three and four explore the relationships between teacher

and peer variables and students' formative achievement, and change in achievement, on criterion-referenced language and literacy assessments that use standardized administration protocols. These assessments are curriculum-based measures of language and literacy, which were validated for use with three to five-year old children. These measures were designed to be sensitive to students' exposure to curriculum and instruction.

3. *Research Question (RQ) 3*: To what extent are selected classroom peer characteristics associated with students' language and literacy achievement and change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?

   H_{03}: Classroom peer characteristics do not affect student language or literacy achievement on progress monitoring measures, or change in student achievement on language or literacy measures

   H_{A3}: Classroom peer characteristics affect student language or literacy achievement on progress monitoring measures, or change in student achievement on language or literacy measures

4. *Research Question (RQ)4*: What is the relationship between teacher characteristics and students' spring language and literacy achievement or fall-to-spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?

   H_{04}: Teacher characteristics do not affect student language or literacy achievement on progress monitoring measures, or change in student achievement on language or literacy measures

   H_{A4}: Teacher characteristics affect student language or literacy achievement on progress monitoring measures, or change in student achievement on language or literacy measures

This is a sample of convenience, as the researcher served as the professional development manager for the DC Partnership for Early Literacy (DCPEL; Partnership), the federally funded Early Reading First grant that supported this work during the 2009-2010 school year. Despite being a sample of convenience, characteristics of this project are aligned with Shadish, Cook and Campbell's main criteria for increasing the quality of study designs that use nonrandom assignment (2002). Specifically, there were multiple dependent variables, and the program implemented a pretest/posttest rather than posttest-only design. These characteristics help to address potential threats to internal validity.

In addition, the data from the sample are ideal for research due to the inclusion of assessments that were nationally normed and standardized on the age group included in the study, or that include standardized administration protocols and established predictive or concurrent validity with norm-referenced measures (Shadish et al., 2002). Finally, the program design includes data collection on key demographic variables (e.g., race/ethnicity, free and reduced price meals [FARM] status, ELL status, and gender) that allow comparison with the normed sample from the norm-referenced assessments, as well as a "secondary data comparison" in which data from this sample might be compared with samples drawn from other studies, such as those analyzed in the preceding literature review (Shadish et al., 2002).

DC Partnership for Early Literacy (DCPEL) Project Components

Curriculum Implementation. DCPEL adopted a comprehensive, outcome-based instructional program based on the results of the meta-analytic study conducted by the National Early Literacy Panel (2004). The program includes goals for children's experiences in the following developmental domains: language, literacy, mathematics,

behavioral, and social-emotional. *Opening the World of Learning* (*OWL*; Schickedanz & Dickinson, 2005) served as the project's core curriculum. *OWL* is a thematic, comprehensive early childhood program with specific academic learning objectives as well as a defined set of daily activities and content to support children's attainment of the objectives. *OWL* aligns with the principles of content centered classroom (Neuman, 2006). In-class coaching was used to help teachers use a variety of data sources to analyze children's strengths and needs and differentiate instruction.

DCPEL Professional Development Program and Activities. DCPEL teachers received ongoing professional development through a mixture of didactic and experiential methods. The core delivery system was through weekly in-class coaching. Teachers were also exposed to formal workshops and trainings throughout the year. Finally, teachers participated in monthly principal-facilitated professional learning communities at their sites. Professional development was provided to all classroom instructional staff (e.g., lead teachers and assistant teachers).

In-Class Coaching. In the 2009-10 school year, DCPEL used Response-to-Intervention Coordinators (RTIC) and Curriculum Coaches to provide ongoing support to teachers in their classrooms on the project's identified professional development foci. Four RTICs had completed master's degrees in early childhood special education (1), special education (1), school psychology (1), and reading specialist (1). The fifth RTIC had a bachelor's degree in English and was completing a master's degree in early childhood special education.

There were two Curriculum Coaches, one with a bachelor's degree in French and the other with a bachelor's degree in liberal arts. Both of the Curriculum Coaches had

demonstrated mastery of the project's core curriculum by working in instructional roles within DCPEL classrooms for three years prior to their promotion to Curriculum Coach. One of the coaches piloted the core curriculum with the curriculum's author. Coaches received ongoing support as part of a professional learning community (PLC) led by the Professional Development Manager. In addition, they received ongoing train the trainer support from nationally recognized expert grant consultants.

The Partnership goal was for each classroom to receive 3-to-4 hours of coaching per week from RTI coaches and Curriculum Coaches, for a total of 108-to-144 hours of coaching throughout the year. Actual hours of coaching received from September to June ranged from 66.3 to 167.0, with a mean of 122.5 hours. Classrooms with higher assessment scores on CLASS and ELLCO received fewer hours of coaching than classrooms with demonstrated need for more coaching.

Professional development workshops. All instructional staff attended a 6-day summer institute in August 2009 that provided training in the *OWL* curriculum, full administration training on CLASS and ELLCO, and scientific findings on language and literacy development and instruction. Site-based and Partnership-wide workshops ranging in length from 1-to-8 hours were held throughout the school year. Classrooms who had participated in year one attended a greater ratio of shorter workshops, whereas newly joining classrooms attended more frequent and longer didactic trainings. Workshop topics were determined based on student achievement and classroom quality data, and DCPEL staff (e.g., curriculum coaches, assessment personnel) and grant consultants developed and led the trainings.

Professional Learning Communities (PLC). Teachers also participated in campus-specific Professional Learning Communities (PLC), in which each group selected books or journal articles from a project-approved list for study and discussion (see Appendix C). PLC meetings were led by the campus principal.

Sample

The data for this study were drawn from the DC Partnership for Early Literacy (DCPEL) preschool and pre-kindergarten classes of 2009-2010. DCPEL is a partnership among six public charter schools across different sections of Washington, DC. The six charter schools work together under the support of a three-year federally funded Early Reading First grant, which began in Fall 2008. In each school there are four to seven classrooms of three and four-year olds. There were roughly 19 children per classroom. Thus, the total number of children participating in the program was 536.

Parent permission forms were provided in English, Spanish, Vietnamese, and French, the major languages spoken and read by DCPEL families. Permission forms informed parents about the language and literacy focus of the federal grant funding, and asked parents for permission to assess their children, share their child's test findings with school staff and in reports and research, and to videotape and photograph children in school settings to support instruction and professional development. In exchange for participation, parents received free books and materials to support their child's learning at home. Two parents did not provide consent for photography and video due to religious beliefs, but all parents gave permission for students to participate in the DCPEL assessment and evaluation.

Teachers also provided consent to participate in federal and academic research and evaluation of DCPEL practices. Teachers completed consent to participate forms, which provided permission for data, videotape and photographs from his or her classroom to be shared for instructional support, professional development, or educational research purposes. One hundred percent of instructional staff (n=68) completed the consent form, which also included a section to collect academic background information. University of Maryland Institutional Review Board (IRB) approved the use of the DCPEL data for this study since parents and teacher gave explicit permission for the data to be used in educational research.

Analytic sample. The full DCPEL database includes data for 536 students. From that total, between 470 and 522 students (87.7% to 97.4%) completed testing for each of the dependent variables. Excluding cases for students missing data from any of the assessments windows retained an analytic sample of 431 students, or 80.4% of students who were participating in the project. Thirty-three children entered after beginning of year, or baseline assessments (6.16%), which occur between the third and sixth week of school. Thirty-nine students withdrew prior to the outcome, or end of year, assessments (7.28%). Preschool and prekindergarten are voluntary grades for District of Columbia residents, so absenteeism and tardiness also contributed to missing data. Multiple imputation (MI) methods were implemented, but evaluation did not demonstrate significant differences between the MI dataset and the analytic sample. Thus, the analytic sample of 431 students was used for the analysis presented in this dissertation.

Program-level comparisons. As indicated by Table 3.1, the analytic sample (n= 431) and full sample (n = 536) demonstrate no statistically significant differences on any

of the demographic variables. Both samples comprise between 55% to 100% African American students across schools. In addition, 77-90% of students across classrooms received free or reduced lunch, a commonly used proxy for parents' income (Harwell, Maeda, and Lee, 2004; Sirin, 2005; Stein et al., 2008).

Finally, between zero and 17% of the total sample's school populations were English language learners (ELL), with fourteen different home languages represented. Spanish, Vietnamese, and French were the most dominant second languages. Male and female students were similarly enrolled. Very few (i.e., less than five percent) entering three-year old students came with an individualized education plan (IEP), and all students who entered with an IEP at this age had speech-language determined as their primary disability category.

Classroom-level comparisons. Included in Table 3.1 are the descriptive data on classroom-level averages for the demographic variables of interest for both the full and analytic samples. T--tests were conducted to determine whether there were significant differences at the classroom level on three demographic variables. Specifically, the average percentage of Hispanic students per classroom with complete assessment data is significantly different between the full sample (5.26%) and analytic sample (4.19%). The analytic sample has a significantly higher percentage of students identified as "Other Race/Ethnicity" per classroom (5.79%) compared to the full sample (4.71%). Despite these classroom-level significant differences for race/ethnicity, there are no significant differences in the average percentage of ELL students per classroom in the full (11.81%) and analytic (12.14%) samples.

For lunch status, the percentage of reduced price lunch students per classroom is significantly lower in the analytic sample (15.98%) compared to the full sample (16.96%), while there is a significant increase in classroom percentage of students who pay full price for lunch in the analytic sample (18.93%) compared to the full sample (17.41%). The classroom average of students who receive free lunch is statistically similar between the full and analytic samples.

Table 3.1. *2009-10 Student Demographics for DCPS, DCPEL and the Analytic Sample*

	DCPS[a] (%)	DCPEL[b] (%)	Analytic (%)	DCPEL vs. Analytic t-value
School Averages		536	431	
3 Years Olds[c]	37.2	53.36	54.01	-0.31
Female	48.8	52.43	52.93	0.21
Race				-0.80
Black	79	85.45	84.07	0.97
Latino	12	5.22	4.45	-0.72
White	7	4.85	5.86	1.14
Other Race/Ethnicity	2	4.48	5.62	
Lunch Status				
Free Lunch	66[d]	65.67	65.34	-0.14
Reduced Price Lunch		16.79	15.46	-0.74
Paid Lunch		17.54	19.20	0.91
ELL	7	11.59	12.18	0.39
Special Education	20	3.73	3.98	0.27
Class Averages				
Female		52.59	53.02	1.03
Race				
Black		85.05	84.22	-1.39
Latino		5.26	4.19	**-2.65**
White		4.98	5.80	-1.68
Other Race/Ethnicity		4.71	5.79	**3.07**
Lunch Status				
Free Lunch		65.63	65.09	-0.46
Reduced Price Lunch		16.96	15.98	**-2.25**
Paid Lunch		17.41	18.93	**2.58**
ELL		11.81	12.14	0.85
SPED		3.85	4.19	0.97
Class Size	mean	18.48	18.48 (14.72[e])	

Note: [a]Data from DCPS; [b]Data from DCPEL partner schools; [c]3-year olds as percentage of total PS/PK enrollment; [d]Aggregate percentage of students receiving free and reduced price lunch ; [e]Average students per class with complete data; Bold data in t-value column indicates statistical significance ($p < 0.05$)

Variables

Information in this section relies heavily on the DCPEL grant narrative and program evaluation documents (AppleTree Institute for Education Innovation, 2007; Ramey, Ramey, Crowell & Polanski, 2009). Table 3.2 lists the names and descriptions for variables used in this research. More information about each of the variables is provided in this section.

Dependent variables. Six language and literacy assessments were administered to all enrolled children at pre-determined assessment points during the 2009-2010 school year:

- baseline - norm-referenced assessments administered from the first through third weeks of September;
- fall - criterion-referenced assessments administered the last two weeks of October;
- winter - criterion-referenced assessments administered the second two weeks of December to all enrolled children;
- spring - criterion-referenced assessments administered the first two weeks of March;
- outcome - norm-referenced assessments administered the last three weeks of May.

Table 3.2. *Names and descriptions for variables used in study.*

	Variable Name	Description
Dependent Variables		
Language	SP_LAN	Spring achievement on IGDI assessment
	FS_LAN	Fall-to-spring change in score on IGDI assessment
	OUT_LAN	Outcome language composite variable derived from EVT-III, PPVT-IV, and TOPEL-DV assessments
Literacy	SP_LIT	Spring language composite variable derived from PALS-LN, PALS-LS, PALS-NW and GRTR assessments
	FS_LIT	Fall-to-spring change in score on literacy composite variables derived from PALS-LN, PALS-LS, PALS-NW and GRTR assessments; difference between fall and spring literacy composite scores
	OUT_LIT	Outcome literacy composite variable derived from TOPEL-PA and TOPEL-PK assessments
INDEPENDENT VARIABLES		
Academic	BASE_LAN	Outcome language composite variable derived from EVT-III, PPVT-IV, and TOPEL-DV assessments
	BASE_LIT	Outcome literacy composite variable derived from TOPEL-PA and TOPEL-PK assessments
	ELL	English language learner status
	TIER3	Special education status
	YR_TWO	Returning students - students in second program year
Demographic	FRP	Child's lunch status - free, reduced or paid
	RACE	Child's race - white, black, Hispanic, or other race
	FEMALE	Child's gender
	FOUR	Child's age at time of baseline assessments
Peer	AVGLIT	Class average of achievement on BASE_LIT
	AVGLAN	Class average of achievement on BASE_LAN
	NELL	Total class number of ELL students
	NTIER3	Total class number of special education students
	NYR_TWO	Total class number of returning students
	NFOUR	Total class number of four-year old students
	NFEMALE	Total class number of female students
	NMIN	Total class number of minority students
	NFRP	Total class number of students receiving free and reduced price lunch
Classroom instruction and quality	LD_DGREE	Lead teacher's highest degree level
	SM_GRP	Number of small groups taught per day
	FS_CL_ES	Fall-to-spring change score on CLASS Emotional Support domain
	FS_CL_CO	Fall-to-spring change score on CLASS Classroom Organization domain
	FS_CL_IS	Fall-to-spring change score on CLASS Instructional Support domain
	FS_EL_CS	Fall-to-spring change score on ELLCO Classroom Structure domain
	FS_EL_CU	Fall-to-spring change score on ELLCO Curriculum domain
	FS_EL_LE	Fall-to-spring change score on ELLCO Language Environment domain
	FS_EL_BB	Fall-to-spring change score on ELLCO Books & Book Reading Opportunities domain
	FS_EL_PE	Fall-to-spring change score on ELLCO Print & Early Writing domain

Data from all but the winter administration are used in this study. These measures are briefly described here and presented in alphabetical order. See Table 3.3 for an outline of all language and literacy student measures presented according to construct.

Table 3.3 also includes the reliability and or validity information for these measures reported in the respective technical manuals. Beginning of year measures are used to analyze and control for any significant differences among children and classrooms at the beginning of the school year. Raw scores on the norm-referenced measures were converted to standardized scores (M = 100, SD = 15) using the conversion tables in each technical manual.

Expressive Vocabulary Test III (EVT-III; Williams, 2007). Children's expressive vocabulary was measured using the EVT-III, a standardized assessment tool (Williams, 2007). On the EVT-III, children complete two main tasks, with administration time averaging about 10 minutes (Williams, 2007). Children label items by naming pictured items, or they describe items by providing synonyms for pictured items. This assessment was administered twice yearly by trained outside assessors as a baseline and outcome measure in September and May.

Get Ready to Read! (GRTR*;* Whitehurst & Lonigan, 2001a). GRTR is a brief 20-item assessment designed to assess children's knowledge of phonological awareness and phonics, book conventions, print and writing aligned with TOPEL measures (Whitehurst & Lonigan, 2001a). Specifically, GRTR is divided into sections: Alphabet Awareness; Print Awareness; Blending & Elision; and Rhyme & Alliteration. Classroom teachers administered this assessment in October, December, and March as a method of progress monitoring using standardized administration. Administration time typically takes about five minutes.

Individual Growth and Development Indicators Picture Naming Subtest *(IGDI-PN;* Missall & McConnell, 2004*)*. The IGDI-PN is a one-minute timed picture-naming

task (Missall & McConnell, 2004). Children are shown pictures one at a time in random order over one minute, and asked to name as many pictures as possible in the allotted time. Classroom teachers trained in standardized administration administered this assessment in October, December, and March as a progress monitoring measure.

Table 3.3. *Student Assessment & Progress Monitoring System*

Construct	Baseline Screening (September & May)	Progress Monitoring (3 times/yr)
Language	*Peabody Picture Vocabulary Test-IV* (Dunn & Dunn, 2007); nationally normed measure of receptive vocabulary; Test-restest *r*=.93 *Expressive Picture Vocabulary Test-III* (Williams, 2007) nationally normed measure of expressive vocabulary; Test-restest *r*=.93 *Test of Preschool Early Language* ([TOPEL] Lonigan et al., 2007), Definitional Vocabulary subtest; measures knowledge of words and their use; Criterion validity=.59 to .77	*Individual Growth and Development Indicators* (Missall & McConnell, 2004); one-minute timed picture naming task; concurrent validity=.56 to .81 (Priest, Davis, McConnell, McEvoy, & Shinn, 1999)
Literacy	*Test of Preschool Early Language* ([TOPEL] Lonigan et al., 2007), Phonological Awareness subtest; measures word elision and blending abilities; Coefficient alpha=87; Test-retest *r*=.83 *TOPEL*, Print Knowledge subtest; measures early knowledge about written language conventions and form; Coefficient alpha=.95; Test-retest *r*=.89	*Get Ready to Read* (Whitehurst & Lonigan, 2001); a brief 20-item assessment designed to assess children's knowledge of phonological awareness and phonics, book conventions, print and writing; Split-half reliability=.80, *PALS-PreK* Name Writing subtest (Invernizzi et al., 2001); measures ability to form letters of name and differentiate written from drawn text; interrater reliability=.99; Criterion validity=.70 to .79 *PALS-PreK* Letter Identification and Letter Sound Identification subtests (Invernizzi et al., 2004); Criterion based letter identification and letter-sound correspondence tasks; interrater reliability=.99

Note: Adapted from 2008 DCPEL Early Reading First Grant Application

PALS Pre-K *(Letter ID, Letter Sounds, Name Writing;* Invernizzi et al., 2001*)*. In the Letter Identification and Letter Sounds subtests, children are asked to identify all letters of the alphabet or 26 sounds including two digraphs, presented in random order on

a standard white sheet of paper. Children's ability to write their own names, and to distinguish their name from a picture, was assessed using the PALS Name Writing subtest. Each child receives a standard sized sheet of white paper and is asked to write their name and draw a self-portrait. Children's work on the Name Writing task is scored according to a seven-point scale. All three subtests were used for baseline, outcome and progress monitoring purposes, in September, October, December, March and May. The PALS Pre-K technical manual describes an approximate administration time of seven to ten minutes for the three assessments combined.

Peabody Picture Vocabulary Test IV *(PPVT-IV*; Dunn & Dunn, 2007*)*. The PPVT-IV measures children's receptive vocabulary. This assessment was administered twice yearly by trained outside assessors as a baseline and outcome measure in September and May. Children are asked to point to one of four pictures after being provided a vocabulary word prompt. Total administration time is estimated at 10-15 minutes (Dunn & Dunn, 2007).

Test of Preschool Early Literacy *(TOPEL*; Lonigan et al., 2007*).* The TOPEL is a standardized measure of three to five year-old children's print knowledge (TOPEL-PK), definitional vocabulary (TOPEL-DV), and phonological awareness (TOPEL-PA) ability. This assessment was administered twice yearly by trained outside assessors as a baseline and outcome measure in September and May. The TOPEL technical manual reports administration time of approximately 30 minutes to complete all three subtests (Lonigan et al., 2007).

TOPEL-PK evaluates children's ability to recognize and understand the use of text and letters/letter sounds within various contexts. TOPEL-DV evaluates children's

ability to identify correctly and then to operationalize familiar objects that they regularly encounter. Finally, TOPEL-PA examines children's ability to associate sounds with specific letters of the alphabet; to subtract sounds or parts of words from words; and to string together letter sounds and parts of words to create whole words. Phonological awareness is an emergent literacy skill that predicts how well children can sound-out words while learning to read.

Final dependent variables. All dependent variables used in this study were recoded to standardized, continuous variables based on sample means and standard deviations prior to analysis (mean = 0; SD = 1). Where possible, language and literacy factor variables were created to reduce the number of dependent variables used in this analysis. Information about the procedures used to create the factors, along with descriptive statistics for each variable, is included in Chapter 4.

Independent variables. Independent variables include instructional quality measures, student demographic and prior achievement data, as well as classroom characteristic data for teachers and peers.

Instructional quality measures. In addition to student measures, fall-to-spring change in ratings on two measures of teacher and classroom quality were analyzed for their predictive role in students' achievement and change in achievement on the language and literacy measures outlined in Table 3.3. Instructional quality assessments were completed using research-driven, standardized classroom observations that probe for multiple factors research has identified as integral to enhancing students' preschool and prekindergarten classroom experiences (Greenberg, Domitrovich, & Bumbarger, 2001; Pianta, 2006). These factors have also been demonstrated as instrumental in fostering

children's capacity to develop strong language and literacy skills (Dickinson & Tabors, 2001).

Classroom Assessment Scoring System (CLASS Pre-K). CLASS Pre-K measures the quality of classroom climate and instructional interactions across ten dimensions in the domains of emotional support, classroom organization, and instructional support (Pianta & La Paro, 2003). See Table 3.4 for the dimensions evaluated under each domain. The observational tool uses a seven point Likert scale, with scores grouped into a low (1, 2), mid (3, 4, 5) or high (6, 7) range. Domain ratings are derived from the average of all dimension ratings during an observation window. The developers reported interrater reliability ranging from 78.8% for Instructional Learning Formats to 96.9% for Productivity. Observers are considered reliable if scores fall within one point of the other on the rating scale.

Table 3.4. *CLASS Domains and Dimensions*

CLASS Domain	CLASS Dimension
Emotional Support (ES)	Positive Climate Negative Climate Teacher Sensitivity Regard for Student Perspectives
Classroom Organization (CO)	Behavior Management Productivity Instructional Learning Formats
Instructional Support (IS)	Concept Development Quality of Feedback Language Modeling

Early Language & Literacy Classroom Observation (ELLCO Pre-K). ELLCO Pre-K measures the quality of classroom climate and instructional interactions for 19 items across five domains: Classroom Structure; Curriculum; Language Environment; Books and Book Reading Opportunities; and Print and Early Writing Supports (Smith, Brady, & Anastasopoulos, 2008). See Table 3.5 for a list of all items, their corresponding domains,

and larger subscales. The observational tool uses a five-point Likert scale. Domain ratings are derived from the average of all item ratings during an observation window, and subscale ratings are derived by averaging the appropriate domain scores for the same observation window. The developers reported Cronbach's alpha of .84 and interrater reliability of 88%. Observers are considered reliable if scores fall within one point of the other on the rating scale.

Table 3.5. *ELLCO Domains and Items*

ELLCO Subscales	ELLCO Domains	ELLCO Items
General Classroom Environment (GCE)	Classroom Structure (CS)	Organization of the Classroom Contents of the Classroom Classroom Management Personnel
	Curriculum (C)	Approaches to Curriculum Opportunities for Child Choice and Initiative Recognizing Diversity in the Classroom
Language & Literacy Subscale (LLS)	Language Environment (LE)	Discourse Climate Opportunities for Extended Conversations Efforts to Build Vocabulary Phonological Awareness
	Books and Book Reading (BB)	Organization of the Book Area Characteristics of Books Books for Learning Approaches to Book Reading Quality of Book Reading
	Print and Early Writing (PE)	Early Writing Environment Support for Children's Writing Environmental Print

CLASS and ELLCO change variables. The two measures, CLASS and ELLCO, were administered over three assessment windows in September, January, and May. Teachers' change in CLASS and ELLCO ratings from the September to May administrations was used to analyze relationships to students' language and literacy achievement or change in achievement on the dependent variables. Specifically, a variable was created that subtracted teachers' fall rating on each CLASS or ELLCO

domain from the respective spring domain rating (CLASS/ELLCO domain change = teacher's spring domain rating - teacher's fall domain rating). Each domain change variable was recoded as a standardized and continuous variable (mean=0, SD=1).

Student characteristics. In addition to the student assessments described previously, other student variables were analyzed as independent variables (See Figure 2). Child demographic data were obtained from school records at the beginning of the school year. Child variables included were gender (female = 0, male = 1); low income as demonstrated by free, reduced-price or paid lunch status (paid= 0, reduced = 1, free= 2), age (three = 0, four = 1); minority (White = 0, Black/Hispanic/Other Ethnicities = 1); disability status (IEP = 0; no IEP = 1); and language status based on established DC Public Schools regulations (ELL = 0, English-only = 1).

Dummy-coded variables were also created for lunch status and race/ethnicity, to be used based on results from initial descriptive analysis. Finally, 68 four-year old students were in their second year of the program, concentrated at three of the six schools. A dummy-coded variable was created to denote returning students (YR_TWO; new=0, returning=1).

Classroom characteristics. Classroom variables include general classroom constructs and teacher variables. Student demographics were also included as classroom variables.

Peer variables. To analyze potential relationships between average classroom demographics and individual student outcomes, variables were created to represent the number of ELL, special education, minority (i.e., Black, Latino, and Other), free or reduced price lunch, female, and returning students per class. Twenty-one classrooms

had no returning students. The percentage of returning students per class for the remaining eight classes ranged from 10.1 to 93.3 percent. These variables were recoded as standardized, continuous variables (mean = 0; SD = 1). Standardized, continuous variables were also created for the average baseline language and literacy achievement in each classroom.

Teacher variables. Teacher demographic variables were obtained through teacher reports and school records, where available. Teachers completed a demographic survey at the beginning of the 2009-10 school year, providing information on highest level of education. A classroom-level variable for lead teacher's degree status was created (0 = bachelor's degree; 1 = master's degree). The remaining teacher variables described in this section might also be used to estimate potential school effects, because variance was primarily related to school-level decisions.

Principals at three of the DCPEL sites required teachers to implement two small groups per day (one reading and one math small group), while the remaining sites implemented one per day as outlined in the core curriculum (one alternating reading or math small group). A dichotomous variable was created to test any relationship between daily number of small group instructional experiences and child language and literacy (0 = one small group daily; 1 = two small groups daily). Small group experiences allow more opportunity for direct interaction between teachers and individual children than do whole group experiences, and 14 out of the 29 DCPEL classrooms instituted two daily small groups during the 2009-10 school year.

An additional classroom variable identified for analysis was teacher-student ratio (0 = greater than 1:6 teacher-student ratio; 1 = 1:6 teacher-student ratio), as classrooms

had varying numbers of students and assigned adults across sites. At three campuses, all classrooms were staffed with three adults 18 students at the beginning of the year. At the remaining campuses, classrooms were staffed with two adults, and classroom size ranged from 11 to 24 students per class, based on space constraints in individual classrooms. A standardized, continuous variable was also created for the number of students per class (mean = 0; SD = 1) to provide additional information, because research demonstrates that class size and teacher-to-student ratio are not synonymous in potential impact on student achievement (Finn & Achilles, 1999; Hanushek, 1999). Analysis of exploratory data demonstrated that both the ratio and the class size variables suffered from multicollinearity when entered together or separately with the remaining variables student-level and peer characteristic data. Therefore, these two variables were not used in the analytic models.

Final dependent and independent variable list. Table 3.2 lists and provides a description for all variables used to answer the research questions.

Procedures

The Early Reading First grant funding the DCPEL provided for external evaluators for the project. The role of the external evaluators is to provide ongoing technical assistance, fidelity monitoring for all program activities, data analysis, and program evaluation throughout the grant award cycle. Programmatic, teacher and student data from each school year within the grant award cycle are used to complete a program evaluation that must be submitted to the US Department of Education annually. In addition, external evaluators monitor the student assessment and teacher observation processes for DCPEL. Specific to the 2009-2010 data of interest in the proposed study,

the evaluators provided ongoing technical assistance and monitoring related to training and administration protocols, and reviewed all raw data for accuracy. Data for this dissertation research were derived from the 2009-2010 DCPEL program evaluation report, which was completed in December 2010.

Student assessment training. DCPEL's Assessment Manager is a developmental psychology doctoral student who has advanced to candidacy and is qualified to teach and train on assessment protocols. The Assessment Manager conducted all trainings for the standardized assessments. In addition, the Assessment Manager conducted initial training on all criterion-referenced assessments. The Assessment Manager held "train the trainer" sessions for all classroom coaches (RTI Coordinators), who in turn conducted refresher trainings for teachers prior to each administration period.

Standardized assessment training. Contracted, independent assessors completed two days of intensive training pertaining to the child assessments analyzed in this study (PPVT-IV, EVT-III, and TOPEL). Training was both didactic and interactive, with prospective assessors being instructed on the administration and scoring of the various psychometric instruments via a PowerPoint presentation, then completing administration practice. As part of the training, the Assessment Manager observed an interactive session that spanned the course of several hours, during which each assessor practiced administering his or her assigned instruments to a partner. The trainer provided continuous, evaluative feedback during this practice session in order to ensure that assessments were being administered and scored with fidelity. Following the training, assessors were assigned to complete specific batteries at various DCPEL campuses.

Assessors who did not demonstrate sufficient reliability on a specific measure were not allowed to conduct the protocol for that measure.

Criterion assessment training. Training for all criterion assessments used in this study (i.e., PALS-PK, GRTR, and IGDIs) was conducted according to the protocol outlined in the technical manual, with the addition of video observations for GRTR and IGDIs - PALS-PK provides an official video to support training for each of its measures. Teachers were initially trained to administer assessments in late August, prior to the September administration of the PALS-PK subtests.

Teacher training consisted of a PowerPoint presentation to overview each assessment protocol; video observation and analysis of each assessment administration; practice administering and scoring the PALS Letter ID and Sounds subtests; and practice scoring and justifying PALS-PK provided samples of the Name Writing subtest. Teachers reviewed and practiced the exact protocols outlined in the measures, and reviewed the frequently asked questions for each subtest. Prior to each of the October, December, March and May criterion assessment administration periods, teachers were trained or refreshed on each of the measures they were responsible for administering during a specific administration period.

Student assessment administration. Administration protocols for the standardized and criterion assessments are described.

Standardized assessment administration. Within three weeks of school entry, in September 2009, and prior to school year's end, in May 2010, trained outside assessors individually administered universal screening and outcome instruments to all students (see Table 3.2). All standardized, norm-referenced assessments were administered twice

yearly as a baseline and outcome measure. Different assessors assessed children on the norm-referenced measures in the baseline and outcome to minimize potential bias. Baseline and outcome child assessments were completed by independent, contracted assessors who had no instructional roles within DCPEL.

The Assessment Manager conducted unannounced reliability observations to sites to monitor and co-score assessment administration. Each contracted assessor was observed during live administration for at least three times per assessment he or she administered (i.e., PPVT-IV, EVT-III, TOPEL). Fidelity to assessment protocols was determined through these unannounced observations, with all assessors demonstrating 0.75 or greater interrater reliability with the Assessment Manager during training and co-scoring. To determine interrater reliability, the Assessment Manager observed assessment administration and recorded student's responses. Both the Assessment Manager and the contract assessor independently scored the assessment. The percentage agreement in scoring was used to derive interrater reliability data.

Criterion assessment administration. Classroom teachers individually administered formal progress monitoring tools three times per year, except for the PALS-PK subtests, which teachers administered at five time points throughout the year. Teachers were provided a one-week window in which to assess all children, with a second make up week factored in for students who were unable to be tested during week one. During the administration periods, teachers provided daily updates to their RTI Coordinators on the number of students assessed.

During the assessment period, RTI Coordinators conducted interrater reliability checks for three students per subtest across all classrooms. RTI Coordinators completed

a standard form to record the teacher's score, her own score, and whether or not there was disagreement. These forms were submitted to the Assessment Manager for analysis at the end of the assessment period. RTI Coordinators provided ongoing training on administration protocols when teacher administration error was observed. Interrater reliability data were not collected for progress monitoring assessments.

Teacher observation training. Training for the teacher quality measures is described.

CLASS observation training. Individuals conducting classroom observations using The Classroom Assessment Scoring System (CLASS) first had to demonstrate sufficient reliability as prescribed by the authors of the CLASS (Pianta, La Paro, & Hamre, 2008). Trainers successfully completed a "train the trainer" session held by CLASS-certified trainers at the University of Virginia prior to implementing observer training. Observers completed a two-day training on the CLASS, whereby they were required to score videotaped, model classrooms in accordance with the CLASS protocol. Observers were required to 1) achieve a score that was 80% reliable with the CLASS master codes overall and 2) assign a score that was within one point of the master coder's score per dimension for at least two out of the five training videos. Data were not retained in a manner that allows analysis of the mean and range for CLASS interrater reliability; however, individual coders were required to demonstrate at least 80% reliability across domains prior to assignment to classrooms.

ELLCO observation training. Individuals conducting classroom observations using the Early Language and Literacy Environment Classroom Observation were trained over the course of one day. Training aligned with prescribed procedures detailed in the

ELLCO training manual. Trainers completed a "train the trainer" session with one of the ELLCO tool's authors prior to implementing training of observers. Observers were exposed to a PowerPoint session that provided an overview detailing the procedures to be followed during an ELLCO observation, as well as video segments to be coded and scored according to ELLCO manual standards. Inter-rater reliability was calculated between coders and was at or above 0.80 for all observers. Data were not retained in a manner that allows analysis of the mean and range for ELLCO interrater reliability; however, individual coders were required to demonstrate at least 80% reliability across domains prior to assignment to classrooms.

Teacher observation implementation. Each observation period included two full observation cycles per classroom, per measure, by four different trained assessors over a two-week period. Two assessors observed and independently coded and scored the same classroom period. The observer pool for both CLASS and ELLCO observations was comprised of a mixture of in-house staff (i.e., employed within some participant facility of DCPEL) and contracted observers. DCPEL employees did not conduct observations at their assigned sites to minimize potential observer bias. In addition, observers were paired in teams of two, and kept running notes, or “scripts”, of their observations during each classroom observation. Scripts were submitted to the Assessment Manager and provided to teachers alongside their ratings for review and professional development.

CLASS observation implementation. Total time spent in classrooms per observation period (i.e., Fall, Winter, Spring) was 120 minutes, equally divided into 2 60-minute observation cycles. Following each 20-minute observation, observers had a 10-

minute window to independently code their scripting and assign individual scores based on evidence collected. A single observation cycle's pattern was:

a. Observe and script for 20 minutes
    i. Pause to code and score for 10 minutes
b. Observe and script for 20 minutes
    i. Pause to code and score for 10 minutes
c. Observe and script for 20 minutes
    i. Pause to code and score for 10 minutes

At the conclusion of this cycle, observer dyads were allotted 30 minutes to discuss their ratings with their observation partner and arrive at a final, combined score for each of the 10 dimensions for each of the three observations. Teachers received one rating for each dimension and domain that averaged the combined scores for the two observation cycles, or six 20-minute observations, within a single observation period.

ELLCO observation implementation. Observers were paired in teams of 2, and completed each ELLCO observation for a period of 60 minutes. These 60 minutes encompassed a 45-minute teacher-child scripted observation, and a 15-minute assessment of the ELLCO-based environmental indicators. At the conclusion of the 60-minute observation cycle, observer teams exited the classroom. Within 30 minutes, each observer arrived at an independent score for each of the 19 indicators based on the evidence collected. Observer dyads then collaborated within an additional 30-minute window in order to concur on a final score for each classroom observed. As with the CLASS, teachers received an aggregate Fall, Winter, and Spring score across ELLCO domains and indicators that averaged the ratings for the two 45-minute observations each

observation period.

Data Analyses

Data analyses included both descriptive and inferential approaches. Prior to descriptive and analytic analyses, analysis of power was estimated, and principal components factor analysis was conducted to reduce the number of dependent variables used in the study.

Power analysis for HLM Models. An a priori power analysis was conducted. Spybrook, Raudenbush, Congdon, and Martinez (2009) describe that in multilevel designs the number of clusters, in this case classrooms ($J = 29$), is more important than the number of participants ($n = 431$). Because the proposed study uses existing data, the number of classrooms is a predefined limitation. *Optimal Design* (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2009) allows a researcher to determine the estimated power that will detect a given effect size with the inclusion of at least one covariate explaining 50% of the variance in outcomes for the model, based on established values in educational research (Bloom, Richburg-Hayes, & Black, 2005), and using default α of 0.05, and variance of 0.10 (Spybrook et al., 2009). Using $J = 29$ classrooms and $n = 19$ students per classroom, *Optimal Design* provided an estimate of the ability to detect an effect size of Cohen's $d = 0.34$ with power of 0.80.

Principal components analysis. Stata 11(StataCorp, 2009), a statistical software program, was used to conduct principal components factor analysis with varimax rotation on the baseline and outcome language and literacy measures, as well as the fall and spring literacy measures. Because one assessment was used for vocabulary during progress monitoring, variable reduction was not needed for fall and spring language

variables. Factors with an eigenvalue greater than one were retained, creating continuous, standardized factor variables (mean = 0, SD = 1). Information about the specific factors created is provided in Chapter 4.

Descriptive statistics. Figure 2 provides a conceptual framework for the model and includes all potential variables to be analyzed. Descriptive statistics were run using Stata 11 (StataCorp, 2009).

Student variables. To analyze student group mean differences in the DCPEL sample based on FARM (free/reduced or paid status), independent sample t-tests were conducted on the continuous dependent variables (e.g., PPVT, TOPEL, EVT) and teacher quality variables (e.g., CLASS IS, ELLCO BB). To check for multicollinearity, zero-order correlations were run to test the relationships between the final dependent variables, and between the classroom instructional and quality variables.

Analytic statistics. Following the exploratory data analysis, I used multiple regression analysis to identify whether or not there was sufficient variance at the group level to justify the use of HLM. Multiple regression provides the percentage of variance in the dependent variable that is explained by one or more independent variables. To mitigate possible violations of test assumptions, robust standard errors were estimated (White, 1980). All dependent variables included in Table 3.2 were included in the multiple regression analysis for each independent variable. Only measures found to be significant for each dependent variable at $p \leq .15$ through multiple regression analysis were retained for the subsequent HLM analysis.

A 2-level hierarchical linear model (HLM; Bryk & Raudenbush, 1992) using HLM 6 (Raudenbush, Bryk, Cheong, & Congdon, 2004) was conducted to answer all

four proposed research questions. HLM was selected for the statistical analysis because it allows for analysis of variance both within and between groups. As the proposed study aims to analyze group effects on individual student language and literacy variables (i.e., students nested within classrooms), HLM is an appropriate choice (Hofmann, 2004). HLM accounts for possible interdependence of student scores for students within the same group, in this case classrooms (Raudenbush & Bryk, 2002). This aligns with Brofenbrenner and Morris's ecological perspective (1998) by taking into account possible moderating effects of the teacher and classroom peers on individual student's achievement.

Using a linear rather than nested model may underestimate standard error, leading to possible overestimation of the magnitude of relationships of interest (Raudenbush & Bryk, 2002). In addition, aggregation bias is a threat when sample characteristics may have different effects at different levels. As an example in the present study, analysis may demonstrate that a student's gender is associated with differential outcomes on a dependent variable. The classroom's gender make-up may explain additional variance on the outcome measure for the individual student. Several independent variables were analyzed at the student and classroom levels, so addressing potential aggregation bias is an important feature of the HLM model.

To examine the relationship between change in quality of teacher-child interactions, or classroom peer characteristics, and student language and literacy achievement, a 2-level HLM was fit. The level-1 model represents the relationship between the student-level dependent variables of interest, which are norm- or criterion-referenced language and literacy measures, and predictor variables (e.g., ELL status,

FARM). Individual student characteristics found significant ($p \leq .15$) in relationship to each of the dependent variables through the multiple regression analyses were entered at level 1 of the HLM. The level-2 model represents the influence of teacher-child interactions or classroom peer characteristics on the dependent variables.

Predictor variables, or covariates, entered at levels 1 and 2 were grand mean centered. This aids the researchers' ability to interpret the coefficients by subtracting the mean value of X (e.g., baseline language achievement) for each student's score (X_{ij}), making the intercept (β_{0j}) the expected mean for the student with a mean score on X_{ij}. Thus, β_{0j} would be the expected outcome language achievement for a student whose baseline achievement equals the mean baseline achievement.

Research questions using norm-referenced dependent measures. For RQ1 and RQ2, the dependent variables used in analysis were the five standardized, norm-referenced measures administered only at baseline and outcome: PPVT, EVT, TOPEL-PK, TOPEL-PA, and TOPEL-DV. As the TOPEL has three subtests that each derives a standard score, all three subtests were treated as individual dependent variables. Independent variables of interest were (a) classroom peer averages on baseline dependent and demographic (i.e., ELL status, special education status, and SES) variables for RQ1 and (b) teacher instruction and quality variables, including fall-to-spring change in teacher-child interaction ratings on CLASS and ELLCO domains for RQ2.

There were $i = 1 \ldots, n_j$ level-1 units nested within $j = 1, \ldots,$ level-2 units (i.e., 431 students nested within 29 classrooms).

Step 1. The unconditional model did not include predictor variables, and replicated a random effects model of 1-way analysis of variance (ANOVA). A different

model was fit for each dependent variable to estimate the grand mean for posttest data. The following equation illustrates the model:

Level 1: $Y_{ij} = b_{0j} + r_{ij}$, where

β_{0j} is the intercept, or the mean dependent variable score for classroom j; and

r_{ij} is the level-1 random effect, or the unique effect of student i in classroom j

Level 2: $b_{0j} = \gamma_{00} + u_{0j}$, where

γ_{00} is the grand mean, or the change in the dependent variable from baseline to outcome; and

u_{0j} is the level-2 random effect, or the unique effect of classroom j

Step 2. At Step 2, the model includes achievement on the dependent variable, controlling for pretest data, and calculated in relationship to child (Level 1) and classroom (Level 2) independent (predictor) variables.

Level 1: $Y_{ijk} = b_{0j} + b_{qj} * X_{qij} + r_{ij}$, where

Y_{ij} is the dependent variable score for student *i* in classroom *j*;

β_{0j} is the intercept, or the mean dependent variable score for classroom j after adjusting for baseline data;

β_{qj} is the slope, or the mean predictor variable score for classroom j;

X_{qij} is the level-1 predictor *n* (i.e., SES, ELL status) for the pretest value for student *i* in classroom *j*; and

r_{ij} is the unique effect of student i in classroom j, or the difference between student i and the group mean after controlling for predictor variables

Level 2: $b_{0j} = \gamma_{00} + u_{0j}$

$b_{pj} = \gamma_{p0}$, where

γ_{00} is the grand mean across classrooms, or the mean change in the dependent variable from time one to time two;

γ_{p0} is the average slope (effect) of the predictor variables on dependent variables across classrooms;

u_{0j} is the level-2 random effect between classrooms on the outcome intercept

Research questions using criterion-referenced dependent measures. To examine the strength of relationship between classroom peer characteristics, or teacher instruction or quality variables, and students' change in achievement and level of achievement on the progress monitoring language and literacy measures, a different HLM was fit. The models mirror those in RQ1 and RQ2, except at Step 1, the model was run for both the change data (e.g., change in IGDI performance from Fall to Spring) and the achievement data (e.g., Spring performance on IGDI) for each dependent variable.

There are five standardized, criterion-referenced progress-monitoring assessments: IGDIs, PALS Letter ID, PALS Letter Sounds, PALS Name Writing, and GRTR. The level-2 model analyzes variation in both change and achievement levels on the progress monitoring assessments as a function of the classroom peer variables and teacher instruction and quality variables.

Summary

Program evaluation data from the 2009-2010 DCPEL participating classrooms was used to explore the relationships between teacher and peer characteristics and

students' language and literacy achievement. Data for 431 three and four-year old children in 29 classrooms was analyzed. Significant level-2 HLM coefficients indicate specific teacher and peer variables that help to predict student language and literacy achievement, or change in achievement, on the dependent variables, after controlling for student-level academic and demographic characteristics.

Chapter 4: Results

The results presented here explore relationships between classroom peer and teacher instruction and quality characteristics and student's language and literacy achievement on both global outcome measures and curriculum-based progress monitoring measures. For a review of variables used in this study, see Table 3.2. Four research questions were analyzed:

1. *Research Question (RQ) 1*: To what extent are classroom peer characteristics associated with students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?
2. *Research Question (RQ) 2*: What is the relationship between teacher characteristics and students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?
3. *Research Question (RQ) 3*: To what extent are classroom peer characteristics associated with students' spring language and literacy achievement or fall-to-spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?
4. *Research Question (RQ) 4*: What is the relationship between teacher characteristics and students' spring language and literacy achievement or fall-to-spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?

Statistical analyses of these questions are reported following preliminary reporting on data preparation and exploration techniques.

This chapter is divided into four sections. First, results are provided for the principal components factor analysis (PCA), conducted to reduce the number of dependent variables used in the study. Section two presents descriptive statistics comparing the analytic sample (n=431) to the DCPEL data (n=536), and reports additional descriptive statistics analyzed to make an initial determination of appropriate variables for the final HLM analysis. The third section of this chapter reports results from the multiple regression analysis for each dependent variable, used to further reduce the number of variables analyzed in the hierarchical linear models (HLM). Finally, I present results from the HLM analyses, which were conducted to analyze between and within classroom variance on the global and curriculum-based dependent variables of interest.

Principal Components Analysis

Principal components analysis using varimax rotation was conducted on dependent variables of interest to potentially reduce the number of dependent variables analyzed using HLM. As all of the assessments administered in the DCPEL project have strong psychometric properties and specific identified constructs (i.e., language or literacy), the individual assessment variables considered to comprise each factor variable were decided a priori.

Table 4.1 provides analytic sample averages for three and four-year old children for all dependent variables prior to the principal components analysis, along with effect sizes for fall to spring growth. Factor variables for baseline and outcome norm-

referenced language and literacy assessments, and for fall and spring (progress monitoring) literacy criterion-referenced assessments were created, according to the constructs measured and validated by the test developers. All variables entered into each factor analysis loaded onto a single factor, with eigenvalues at or above 2 for each factor (1.96 to 3.06). Table 4.2 reports the factor loading coefficients for each variable loaded during a PCA.

Table 4.1. *Analytic Sample 2009-10 Averages for Dependent Variables*

	N	Min	Max	*M*	*SD*	Effect Size *Cohen's d*
LANGUAGE MEASURES						
Norm-referenced summative measures (standard scores)						
PPVT-IV						
3-yo Baseline	232	26	140	95.1	15.06	
3-yo Outcome	232	34	136	100.8	15.07	0.38
4-yo Baseline	199	61	144	97.27	14.42	
4-yo Outcome	199	70	141	101.68	13.34	0.32
Total Baseline	431	26	144	96.1	14.79	
Total Outcome	431	34	141	101.2	14.29	0.35
EVT-IV						
3-yo Baseline	232	64	137	99.65	12.73	
3-yo Outcome	232	54	147	104.81	13.04	0.40
4-yo Baseline	199	68	150	99.64	13.88	
4-yo Outcome	199	34	149	102.51	15.12	0.20
Total Baseline	431	64	150	99.65	13.26	
Total Outcome	431	34	149	103.74	14.07	0.30
TOPEL-DV						
3-yo Baseline	232	57	139	85.36	13.48	
3-yo Outcome	232	55	125	99.07	13.87	1.00
4-yo Baseline	199	58	125	94.48	13.54	
4-yo Outcome	199	9	127	100.39	15.01	0.41
Total Baseline	431	57	139	89.57	14.24	
Total Outcome	431	9	127	99.68	14.41	0.71
Criterion-referenced formative measures (raw scores)						
IGDI-PN						
3-yo Fall	232	1	37	15.71	6.21	
3-yo Spring	232	6	39	20.76	6.11	0.82
4-yo Fall	199	0	38	20.1	6.38	
4-yo Spring	199	4	46	23.78	6.59	0.57
Total Baseline	431	0	38	17.74	6.66	
Total Spring	431	4	46	22.16	6.51	0.67

Table 4.1. cont. *Analytic Sample 2009-10 Averages for Dependent Variables*

	N	Min	Max	*M*	*SD*	Effect Size *Cohen's d*
LITERACY MEASURES						
Norm-referenced summative measures (standard scores)						
TOPEL-PA						
3-yo Baseline	232	63	144	89.26	12.71	
3-yo Outcome	232	45	136	95.26	14.42	0.44
4-yo Baseline	199	25	129	91.93	15.99	
4-yo Outcome	199	37	131	100.82	18.15	0.52
Total Baseline	431	25	144	90.49	14.36	
Total Outcome	431	37	136	97.83	16.46	0.48
TOPEL-PK						
3-yo Baseline	232	67	144	100.53	16.91	
3-yo Outcome	232	79	144	117.72	15.42	1.06
4-yo Baseline	199	73	133	105.69	15.22	
4-yo Outcome	199	71	129	112.53	9.15	0.56
Total Baseline	431	67	144	102.91	431	
Total Outcome	431	71	144	115.33	431	0.84
Criterion-referenced formative measures (raw scores)						
PALS-LN						
3-yo Fall	232	0	26	11.15	9.49	
3-yo Spring	232	0	26	19.02	7.97	0.90
4-yo Fall	199	0	26	18.83	8.2	
4-yo Spring	199	2	26	23.46	5.05	0.70
Total Baseline	431	0	26	14.69	9.69	
Total Spring	431	0	26	21.07	7.13	0.75
PALS-LS						
3-yo Fall	232	0	24	4.24	5.87	
3-yo Spring	232	0	25	10.42	7.25	0.94
4-yo Fall	199	0	25	12.04	7.25	
4-yo Spring	199	0	26	19.06	6.35	1.03
Total Baseline	431	0	25	7.84	7.61	
Total Spring	431	0	26	14.41	8.09	0.84
PALS-NW						
3-yo Fall	232	0	7	3.2	1.76	
3-yo Spring	232	0	7	4.62	1.9	0.78
4-yo Fall	199	0	7	5.43	1.86	
4-yo Spring	199	0	7	6.35	1.22	0.60
Total Baseline	431	0	7	4.23	2.12	
Total Spring	431	0	7	5.42	1.84	0.62
GRTR						
3-yo Fall	232	0	20	10.36	3.89	
3-yo Spring	232	4	20	14.52	3.39	1.14
4-yo Fall	199	0	20	15.24	3.78	
4-yo Spring	199	4	20	17.56	2.74	0.71
Total Baseline	431	0	20	12.61	431	
Total Spring	431	4	20	15.92	431	0.82

Language factors. To reduce the number of baseline and outcome dependent variables that assess language, a principal components analysis was conducted using the

EVT-III, PPVT-IV, and TOPEL-DV variables. All language assessments demonstrate strong psychometric properties for the language construct. Baseline EVT-III, PPVT-IV, and TOPEL-DV scores were loaded to create BASE_LAN (baseline language factor variable). Outcome EVT-III, PPVT-IV, and TOPEL-DV scores were loaded to create OUT_LAN (outcome language factor variable). Fall and spring language data were obtained from a single assessment measure (IGDI-Picture Naming), so a factor variable could not be created from the criterion-referenced language data.

Literacy factors. Variables were combined for baseline, fall, spring, and outcome literacy assessments to potentially decrease the number of dependent literacy variables used in later analyses. For baseline (Base_LIT) and outcome (OUT_LIT) literacy factor variables, baseline and outcome TOPEL-PA and TOPEL-PK were loaded, respectively. For fall (FA_LIT), spring (SP_LIT) and fall-to-spring change (FS_LIT) literacy factors, fall, spring and fall-to-spring change variables for GRTR, PALS-LN, PALS-LS, and PALS-NW were loaded in the appropriate PCA.

Table 4.2. *Factor Loadings for Principal Components Analysis of DCPEL 09-10 Student Achievement Data*

Norm-Referenced Language Variables	Baseline Language Factor	Outcome Language Factor	
EVT III	0.904	0.841	
PPVT-IV	0.875	0.858	
TOPEL-DV	0.832	0.788	
Coefficient alpha	0.838	0.772	
Norm-Referenced Literacy Variables	Baseline Literacy Factor	Outcome Literacy Factor	
TOPEL-PA	0.827	0.840	
TOPEL-PK	0.827	0.840	
Coefficent alpha	0.778	0.751	
Criterion Referenced Literacy Variables	Fall Literacy Factor	Spring Literacy Factor	Fall-Spring Literacy Change Factor
GRTR	0.888	0.887	0.606
PALS-LN	0.896	0.887	0.807
PALS-LS	0.899	0.916	0.672
PALS-NW	0.809	0.787	0.698
Coefficient alpha	0.826	0.820	0.589

The language and literacy factor variables were created as standardized, continuous variables (mean = 0, SD = 1). These factor variables were used to conduct analyses reported in the remaining sections of this chapter.

Data Exploration

Student achievement data. The results of the data analysis for the analytic sample of students from the DCPEL program demonstrate strong pre-/post gains, demonstrating program effectiveness. Three and four-year old children in the program made significant gains across standardized and curriculum-based language and literacy measures (Table 4.1). Total effect sizes for standardized measures ranged from 0.30 SD (EVT-III) to 0.84 SD (TOPEL-PK).

Three-year olds' effect sizes on standardized language measures ranged from 0.38 SD on the PPVT-IV to 1.00 SD on the TOPEL-DV. For the standardized literacy measures, three-year old students demonstrated gains of 0.44 SD on the TOPEL-PA and 1.06 SD on TOPEL-PK. Three-year old effect sizes on the curriculum-based measures were all large. Children made 0.82 SD gains on the language measure, and 0.78 SD (PALS-NW) to 1.14 SD (GRTR) gains on the progress monitoring literacy measures.

Gains for four-year old students were more modest but still significant, and included 34.2 percent of students in their second program year. Four-year olds' effect sizes on standardized language measures ranged from 0.32 SD on the PPVT-IV to 0.41 SD on the TOPEL-DV. For the standardized literacy measures, four-year old students demonstrated gains of 0.52 SD on the TOPEL-PA and 0.56 SD on TOPEL-PK. Four-year old effect sizes on the curriculum-based measures were all medium. Children made 0.57 SD gains on the language measure, and 0.60 SD (PALS-NW) to 1.03SD (PALS-LS)

gains on the progress monitoring literacy measures

Table 4.3 demonstrates there were no significant differences in baseline achievement between the full and analytic samples for any of the dependent variables.

Table 4.3. *2009-10 Dependent Variable Averages for DCPEL and the Analytic Sample*

	DCPEL (SD)	n	Analytic Sample (SD)	n	DCPEL vs. Analytic t-value
LANGUAGE					
Baseline Language (standard)	94.89 (12.34)	486	95.08 (12.30)	431	-0.34
Fall Language (raw)	17.32 (6.82)	519	17.77 (6.66)	431	-1.54
Spring Language (raw)	21.82 (6.72)	496	22.14 (6.55)	431	-1.09
Outcome Language (standard)	101.12 (12.32)	470	101.55 (11.86)	431	-0.79
LITERACY					
Baseline Literacy (standard)	95.70 (12.79)	494	96.73 (12.63)	431	-1.81
Fall Literacy (raw)	10.78 (5.56)	522	11.21 (5.47)	431	-1.80
Spring Literacy (raw)	15.42 (4.94)	505	15.78 (4.73)	431	-1.71
Outcome Literacy (standard)	105.99 (12.87)	476	106.71 (12.22)	431	-1.29

***p≤.001, **p≤.01, *p≤.05

Student achievement related to lunch status. Analysis of t-test values for students' performance on the two control and six dependent variables used in this research demonstrated statistically significant differences in performance based on lunch status (Table 4.4). Students who received free or reduced price lunch scored less well than their peers whose families paid full price for lunch on the norm-referenced language and literacy assessments at baseline and outcome (beginning and end of school year). Lunch status was also related to spring literacy achievement on the criterion assessments, with children whose families paid for lunch scoring higher than children receiving free or reduced price lunch, although children who received free or reduced price lunch made more fall-to-spring gains on the same measures than their paid peers. There were no significant differences between children's lunch status and spring performance, or change in performance from fall to spring, on the criterion language measure.

Table 4.4. *Descriptive data exploring differences in performance on norm- and criterion-referenced assessments by DCPEL children whose families paid or received free or reduced price lunch during the 2009-10 school year*

	Paid	Free/Reduced	t-value
Control variables			
BASE_LAN	0.61	-0.12	6.24***
BASE_LIT	0.52	-0.24	4.54***
Dependent variables			
FS_LAN	-0.01	-0.02	0.14
SP_LAN	0.17	0.00	1.40
OUT_LAN	0.64	-0.10	6.62*
FS_LIT	-0.17	0.05	-1.77*
SP_LIT	0.35	0.02	2.80**
OUT_LIT	0.54	-0.07	5.34***

***p≤.001, **p≤.01, *p≤.05

Quantile-Quantile (Q-Q) plots were observed for baseline to outcome, fall to spring, and fall to fall-to-spring change variables for language and literacy. Graphs demonstrated relatively normal distributions for all variables, although fall to spring and fall to fall-to-spring literacy observations displayed slight deviance from the fitted lines at the lower ends.

Zero-order correlation matrices were created to evaluate potential multicollinearity among the dependent variables, and the classroom instruction and quality variables (see Table 4.5). Correlation values greater than 0.80 can indicate collinearity concerns. As another step in preliminary data analysis, variance inflation factors (VIF) values were run for each of the variables. VIF values provide an estimate of how much the variance of a single coefficient estimate is inflated by multicollinearity. VIF values greater than 10 indicate serious concerns with multicollinearity (Kutner, 2004). Analysis of the correlation matrices demonstrates that correlations on the dependent variables range from -0.06 to 0.65, and VIF values are between 1.15 and 2.32. These numbers do not indicate possible multicollinearity concerns, as data are

substantially below the accepted range (Kutner, 2004). All dependent variables were retained for further analysis.

Table 4.5. *Correlations and variance inflation factors between dependent variables in the analytic sample*

Dependent variables	VIF value	out_lan	out_lit	sp_lit	fs_lit	fs_lan	sp_lan
OUT_LAN	2.05	1.00					
OUT_LIT	2.32	0.65	1.00				
SP_LIT	2.05	0.42	0.59	1.00			
FS_LIT	1.15	0.01	0.14	0.29	1.00		
FS_LAN	1.62	-0.06	-0.05	-0.08	0.06	1.00	
SP_LAN	2.04	0.35	0.20	0.37	0.01	0.50	1.00

Classroom quality data. Table 4.6 and 4.7 report the mean, standard deviation, and range of ratings on the CLASS domains ELLCO subscales and domains for the 29 DCPEL classrooms across fall, winter and spring observation windows. Teachers in DCPEL classrooms made significant gains across quality measures from fall to spring 2009-10. Ratings for fall, winter and spring on all CLASS domains demonstrate higher than reported means for pre-kindergarten classrooms at fall and spring (Mashburn et al., 2008). There are no published studies using the revised ELLCO (Smith, Brady, & Anastasopoulos, 2008).

Table 4.6. *DCPEL 2009-10 Averages Across CLASS Domains*

	FALL				WINTER				SPRING				FALL to SPRING Change
CLASS DOMAIN	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	Cohen's *d*
ES	5.49	0.68	3.63	6.75	5.60	0.71	3.90	6.75	5.78	0.47	4.74	6.63	.50
CO	5.08	0.86	3.22	6.67	5.23	0.90	3.40	6.37	5.37	0.62	3.78	6.57	.38
IS	3.04	0.65	1.95	4.67	3.75	1.03	1.60	5.30	3.95	0.72	2.67	5.72	1.32

Note. Emotional Support (ES); Classroom Organization (CO); and Instructional Support (IS)

Table 4.7. *DCPEL 2009-10 Averages Across ELLCO Domains and Subscales*

ELLCO Subscales & Domains	FALL				WINTER				SPRING				FALL to SPRING Change
	M	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	Cohen's *d*
GCE	3.76	0.53	2.57	4.71	4.07	0.49	2.82	5.00	4.28	0.44	3.07	5.00	1.06
CS	4.06	0.50	2.63	4.75	4.26	0.47	3.13	5.00	4.46	0.41	3.13	5.00	.91
CU	3.88	0.58	2.50	4.67	3.88	0.58	2.50	5.00	4.09	0.54	2.83	5.00	1.05
LLS	3.33	0.65	2.16	4.69	3.78	0.54	2.45	4.73	4.10	0.48	2.58	4.78	1.35
LE	3.22	0.75	2.00	4.75	3.57	0.68	1.75	4.50	3.85	0.56	2.25	4.75	.99
BB	3.50	0.72	1.80	5.00	4.07	0.43	3.10	4.70	4.22	0.49	3.00	5.00	1.21
PE	3.34	0.69	2.00	4.67	3.81	0.67	2.50	5.00	4.22	0.53	2.50	5.00	1.47

Note. General Classroom Environment Subscale (GCE), Classroom Structure (CS), Curriculum (CU); Language & Literacy Subscale (LLS), Language Environment (LE), Books & Book Reading Opportunities (BB), and Print & Early Writing (PE)

Analysis of t-test values for CLASS and ELLCO data (Table 4.8) demonstrated that students whose families paid full price for lunch were in classrooms with higher average CLASS Instructional Quality, and higher average ratings on all of the ELLCO domains. There were no significant differences between classroom quality ratings for CLASS Emotional Support and Classroom Organization for classrooms of students receiving free or reduced price lunch when compared to classrooms of students whose families paid full price for lunch.

Table 4.8. *Descriptive data exploring differences in average instructional quality experienced by DCPEL children whose families paid or received free or reduced price lunch during the 2009-10 school year*

	Paid	Free/Reduced	t-value
CL_ES	5.65	5.60	0.87
CL_CO	5.27	5.16	1.45
CL_IS	3.71	3.55	2.14*
EL_CS	4.33	4.25	1.90*
EL_CU	3.95	3.78	2.89**
EL_LE	3.66	3.52	2.17*
EL_BB	4.03	3.85	3.31***
EL_PE	3.91	3.76	2.34**

***p≤.001, **p≤.01, *p≤.05

When all of the classroom peer variables were entered with the teacher instruction and quality variables to test for variance inflation, potential issues of multicollinearity surfaced. Beginning with number of ELL (NELL), and number of special education students (NTIER3) as the core peer predictor variables, additional predictor variables were entered and tested to check for multicollinearity. The minimum VIF value for teacher-to-student ratio was 12.68, and the minimum value for number of students per class was 10.28 regardless of configuration of variables. This is likely due to limited variance in class size and teacher-to-student ratio. When these two variables were removed from the collinearity diagnostics, the maximum VIF for the classroom-level variables was 5.95 (NFRP), and the VIF value for retained variables ranged from 1.29 to 5.95. Correlations for the retained variables ranged from -0.50 to 0.82 (Table 4.9). These values are acceptable to reduce possible multicollinearity concerns (Kutner, 2004), so variables were retained for further analysis.

Table 4.9. *Correlations and variance inflation factors between classroom peer, instruction and quality variables in the analytic sample*

	VIF Value	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. NMIN	4.38	1.00														
2. NELL	5.00	-.36	1.00													
3. NTIER3	5.92	-.17	.33	1.00												
4. NYR_TWO	4.80	.05	.17	.79	1.00											
5. NFRP	5.95	.82	-.49	-.12	.14	1.00										
6. LD_DGREE	1.66	.35	-.31	.02	.07	.47	1.00									
7. SM_GRP	5.38	.45	-.78	-.53	-.39	.50	.24	1.00								
8. FS_CL_ES	3.10	.15	.10	-.19	-.15	.25	.36	.06	1.00							
9. FS_CL_CO	3.52	.29	.10	-.17	-.12	.34	.25	.06	.72	1.00						
10. FS_CL_IS	2.14	.12	-.14	-.50	-.26	.09	.13	.17	.22	.34	1.00					
11. FS_EL_CS	3.52	.16	.05	-.46	-.36	.12	.05	.29	.20	.12	.40	1.00				
12. FS_EL_CU	2.97	.08	-.04	-.27	-.06	.02	-.19	.18	-.23	-.05	.24	.47	1.00			
13. FS_EL_LE	2.87	.07	-.02	-.31	-.29	.11	.07	.33	.22	.14	.24	.71	.46	1.00		
14. FS_EL_BB	4.24	.07	.10	-.34	-.20	.16	-.01	.21	.31	.40	.38	.69	.56	.69	1.00	
15. FS_EL_PE	2.58	.27	-.10	-.38	-.29	.12	-.04	.41	.03	.08	.22	.55	.62	.53	.52	1.00

Multiple Regression (MR) Analysis

As a final method to reduce the number of potential control variables prior to HLM analysis, multiple regression analyses were performed for each of the six dependent variables. This final step allowed for elimination of non-significant student, peer, and classroom instruction and quality control variables. All independent variables were entered into the model for each dependent variable. In addition, the progress monitoring dependent variables were entered into the OUT_LAN and OUT_LIT models, to determine if children's spring achievement on the progress monitoring measures, or change in achievement from fall to spring, predicted outcome language or literacy achievement beyond any relationship to students' baseline achievement. Post-hoc power analysis demonstrated that $R^2\Delta$=0.03 could be detected for 30 variables with 0.95 power at α=0.05, minimal n=272. The analytic sample size of 431 students and number of dependent variables (see Table 3.5) met these criteria.

Tables 4.10 through 4.15 report statistically significant predictors from the multiple regression analysis ($p \leq .15$). Data in each table are organized by student, peer and classroom instruction and quality variables that demonstrated a significant relationship to the dependent variable entered for that model. Factors found to be significant in the MR tables were retained in the subsequent HLM analyses. All six of the regression models significantly predicted child language or literacy outcomes, with F(67.85.193) = .000 to F(264.646) = .000. R^2 for the models ranged from .156 to .631, and the adjusted R^2 from .099 to .604. The unstandardized regression coefficients (B) and standardized regression coefficients (β) are summarized in the tables.

Norm-referenced results of the MR analysis. Tables 4.10 and 4.11 report the significant predictor variables for the standardized, norm-referenced end-of-year measures.

Student variables. Baseline and spring language or literacy achievement was significantly and positively predictive of students' year-end achievement. However, change in language achievement from fall to spring was negatively related to children's language outcomes. This may be because students who experienced more fall-to-spring gains also began lower than their peers who demonstrated less gain. Four-year old students significantly outperformed their three-year old peers on the outcome language measure (0.33 SD), but performed less well than their three-year old peers on the outcome literacy measure (-0.59 SD).

ELL student status was positively related to outcome language. Gender was a significant predictor for outcome literacy, with females slightly outperforming males (0.09 SD). Children in their second year of the program did not perform as well as their first year peers (-0.26 SD).

A student's race did not significantly predict either outcome language or literacy achievement. Students who received free lunch scored less well than their peers whose families paid reduced or full price for lunch on the outcome language measure (-0.19 SD). Students who received free and reduced price lunch scored less well than their paid peers on the outcome literacy variable (-0.28 SD and -0.37 SD, respectively).

Peer variables. While an individual student's race did not demonstrate significant relationship to language and literacy achievement, classroom peer composition of minority students did demonstrate a significant and negative relationship for outcome

language and outcome literacy, whereby a one standard deviation increase in number of minority students predicted a -0.12 SD difference in outcome language and literacy for individual students. The number of ELL students per class was a positive predictor of students' language outcomes (0.12 SD). Finally, the number of students receiving free and reduced price meals per classroom was positively related to students' end-of-year literacy achievement (0.16 SD).

***Classroom instruction and quality variables*.** Students whose lead teachers held a master's degree significantly outperformed their peers in classrooms where teachers held a bachelor's degree. On the language composite, lead teacher's degree status predicted a 0.13 SD increase in outcome achievement. On the literacy composite, degree status predicted a 0.27 SD increase.

CLASS domain predictors. Teachers' change in rating from fall to spring on the CLASS Instructional Support domain (FS_CL_IS) was also a positive predictor of students' outcome achievement, with a one standard deviation gain in rating predicting a 0.14 SD difference in outcome language achievement and a 0.13 SD difference in outcome literacy achievement. However, teachers' fall-to-spring change in rating on the CLASS Classroom Organization (FS_CL_CO) domain was negatively related to students' literacy achievement. Every one standard deviation gain in rating on the domain predicted a -0.19 SD difference in students' outcome literacy achievement.

ELLCO domain predictors. Teachers' change in rating on one domain of the ELLCO was found to significantly predict students' performance related to the outcome assessments. Teachers' fall-to-spring change on the ELLCO Books & Book Reading

domain (FS_EL_BB) was negatively related to outcome achievement in language (-0.10 SD).

Table 4.10. *Summary of Multiple Regression Analysis for Variables Predicting Outcome Language* (OUT_LAN)

Variable	B	β
Student Variables		
BASE_LAN	0.641***	0.666
SP_LAN	0.125***	0.130
FS_LAN	-0.181***	-0.190
FOUR	0.328**	0.112
ELL	0.328**	0.112
FREE	-0.187**	-0.093
YR_TWO	-0.259**	-0.099
Peer Variables		
NMIN	-0.115**	-0.122
NELL	0.116**	0.121
Classroom Instruction & Quality Variables		
LD_DGREE	0.126*	0.063
FS_CL_IS	0.139**	0.144
FS_EL_BB	-0.099*	-0.103

***p≤.05, **p≤.10,*p≤.15

Table 4.11. *Summary of Multiple Regression Analysis for Variables Predicting Outcome Literacy* (OUT_LIT)

Variable	B	β
Student Variables		
BASE_LIT	0.340***	0.347
SP_LIT	0.564***	0.556
FOUR	-0.592***	-0.305
FEMALE	0.089*	0.046
FREE	-0.282***	-0.139
REDUCED	-0.368***	-0.138
Peer Variables		
NMIN	-0.115**	-0.121
NFRP	0.156**	0.165
Classroom Instruction & Quality Variables		
LD_DGREE	0.270***	0.134
FS_CL_CO	-0.188***	-0.186
FS_CL_IS	0.125***	0.128

Criterion results of the multiple regression analysis. Tables 4.12 through 4.15 report the significant predictor variables for the criterion-referenced progress monitoring measures, and the fall-to-spring change scores for those measures.

Student variables. For the spring dependent variables, both baseline performance on the global measures and fall-to-spring change in performance on the curriculum-based measures were entered as predictors. Baseline and fall-to-spring change in language or literacy achievement was significantly and positively predictive of students' spring language and literacy achievement (SP_LAN; SP_LIT). For language, change in achievement from fall to spring was more strongly related to students' spring language achievement than was baseline language achievement. Analyzing the spring literacy model demonstrates that baseline literacy achievement on the standardized assessments was a stronger predictor of spring achievement than fall-to-spring change on the progress monitoring measures. Baseline literacy predicted lower gains for students

from fall to spring on the literacy measures (FS_LIT). There was no significant relationship between baseline language scores and students' change in language achievement from fall to spring on the progress monitoring measures (FS_LAN).

Special education status (TIER3) was positively related to fall-to-spring change in language achievement (0.62 SD), while students' special education status was negatively related to spring literacy achievement (-0.25 SD). Age was a significant predictor for spring literacy achievement, with four-year olds outperforming three-year olds (0.51 SD). Children who received free lunch outperformed their peers whose families paid reduced or full price for lunch on the spring language measure (0.18 SD), but performed less well than their peers on the spring literacy variable (-0.16 SD).

Peer variables. A classroom's average baseline achievement on the standardized assessments predicted fall-to-spring change on the language progress monitoring measure, with a one standard deviation increase in average baseline language achievement predicting 0.15 SD lower gains in language achievement from fall to spring. Although an individual student's race did not demonstrate significant relationship to spring language and literacy achievement, or change in achievement from fall to spring, classroom peer composition of minority students did demonstrate a significant and positive relationship to spring language (0.26 SD). Composition of minority students demonstrated a negative relationship to fall-to-spring change in achievement for language and literacy.

Composition of ELL students demonstrated a positive relationship to fall-to-spring change in language and literacy achievement, but a negative relationship to spring language achievement. This suggests that students in classrooms with higher numbers of

ELL students were able to catch up with their English-only peers on the literacy measures but not the language measures. Classroom composition of students receiving free and reduced price lunch was negatively associated with spring language achievement but positively related to change in language achievement from fall to spring. This relationship mirrors the relationships demonstrated for composition of ELL and minority students. Finally, the number of second year students was positively related to spring language achievement but negatively related to spring literacy achievement. NYR_TWO was the only peer predictor variable found to be significant in the spring literacy model.

Classroom instruction and quality variables. Students whose lead teachers held a master's degree significantly outperformed their peers in classrooms where teachers held a bachelor's degree on the spring language measure (0.31 SD). Students in classrooms that taught two small groups per day made significantly higher gains in language and literacy achievement from fall to spring than their peers in classrooms having one small group per day (0.83 SD and 0.50 SD, respectively).

CLASS domain predictors. Teachers' change in rating from fall to spring on the CLASS Instructional Support domain (FS_CL_IS) was a negative predictor of spring language (-0.16 SD), but a positive predictor of students' literacy achievement, (0.17 SD), and fall-to-spring change in language and literacy achievement (0.16 SD and 0.21 SD, respectively). Teachers' change in rating from fall to spring on the Emotional Support domain (FS_CL_ES) was negatively related to students' spring language achievement (-0.15 SD), and teachers' change on the Classroom Organization domain (FS_CL_CO) was negatively related to spring literacy achievement (-0.10 SD). Teachers' fall-to-spring change in rating on the CLASS Classroom Organization (FS_CL_CO) domain was

negatively related to students' literacy achievement. Every one standard deviation gain in rating on the domain predicted a -0.10 SD difference in students' spring literacy achievement. This same relationship was demonstrated between FS_CL_CO and outcome literacy, as well.

ELLCO domain predictors. ELLCO domains were much more predictive of progress monitoring assessments than for outcome assessments, although positive and negative relationships were demonstrated. Two ELLCO domains predicted spring language achievement, three domains predicted change in language achievement from fall to spring, and one domain predicted change in literacy achievement from fall to spring.

Teachers' fall-to-spring change on the ELLCO Curriculum domain (FS_EL_CU) was negatively related to spring language achievement (-0.17 SD), but positively related to change in language achievement from fall to spring (0.31 SD). Unlike the negative relationship demonstrated between teachers' change in rating on FS_EL_BB and students' outcome language achievement, students' spring language achievement was positively related to students' spring language achievement (0.14 SD). However, teachers' increased ratings on FS_EL_BB was negatively related to students' change in language achievement from fall to spring (-0.30 SD).

Teachers' rating improvement on the ELLCO Classroom Structure domain was positively related to students' change in language performance from fall to spring (0.26 SD). The final significant ELLCO domain, Print and Early Writing (FS_EL_PE) was negatively related to change in language achievement from fall to spring (-0.43 SD) but positively related to change in literacy achievement for the same timeframe (0.11 SD).

Table 4.12. *Summary of Multiple Regression Analysis for Variables Predicting Spring Language* (SP_LAN)

Variable	B	β
Student Variables		
BASE_LAN	0.392***	0.393
FS_LAN	0.493***	0.500
FOUR	0.508***	0.256
FREE	0.175**	0.084
Peer Variables		
NMIN	0.263***	0.269
NFRP	-0.480***	-0.496
NELL	-0.200***	-0.201
NYR_TWO	0.136**	0.138
NTIER3	-0.145**	-0.146
Classroom Instruction & Quality Variables		
LD_DGREE	0.305***	0.148
FS_CL_ES	-0.152**	-0.146
FS_CL_IS	-0.159***	-0.159
FS_EL_CU	-0.172***	-0.173
FS_EL_BB	0.139**	0.138

Table 4.13. *Summary of Multiple Regression Analysis for Variables Predicting Fall-to-Spring Language Change* (FS_LAN)

Variable	B	β
Student Variables		
BASE_LAN	0.081*	0.053
TIER3	0.620***	0.120
Peer Variables		
AVGLAN	-0.153***	-0.155
NMIN	-0.301***	-0.304
NFRP	0.205**	0.209
NELL	0.196**	0.194
Classroom Instruction & Quality Variables		
SM_GRP	0.829***	0.392
FS_CL_IS	0.156***	0.154
FS_EL_CS	0.264***	0.262
FS_EL_CU	0.308***	0.306
FS_EL_BB	-0.302***	-0.297
FS_EL_PE	-0.425***	-0.419

***p≤.05, **p≤.10,*p≤.15

Table 4.14. *Summary of Multiple Regression Analysis for Variables Predicting Spring Literacy* (SP_LIT)

Variable	B	β
Student Variables		
BASE_LIT	0.515***	0.533
FS_LIT	0.306***	0.323
TIER3	-0.253*	-0.054
FREE	-0.163**	-0.081
Peer Variables		
NYR_TWO	-0.139**	-0.146
Classroom Instruction & Quality Variables		
FS_CL_CO	-0.104*	-0.104
FS_CL_IS	0.174***	0.181

***p≤.05, **p≤.10,*p≤.15

Table 4.15. *Summary of Multiple Regression Analysis for Variables Predicting Fall-to-Spring Literacy Change* (FS_LIT)

Variable	B	β
Student Variables		
BASE_LIT	-0.367***	-0.359
Peer Variables		
NMIN	-0.265***	-0.267
NELL	0.149*	0.147
Classroom Instruction & Quality Variables		
SM_GRP	0.501***	0.236
FS_CL_IS	0.211***	0.207
FS_EL_PE	0.108*	0.106

Summary of MR analyses. For each dependent variable, at least one predictor variable was demonstrated as significant at the student and classroom levels. Below is a summary of predictor variables demonstrated as significant for both standardized norm-referenced assessments and curriculum-based criterion assessments.

Student predictors. Baseline language or literacy achievement was predictive of achievement or change in achievement for all dependent variables except fall to spring

language achievement. Change in fall-to-spring language and literacy achievement helped to positively predict outcome and spring language and literacy achievement. Student demographic variables were differentially significant across dependent variables. Lunch status was significant for four dependent variables, age was a significant predictor for three, special education status for two, and gender, ELL status and year in program were each significant predictors for one dependent variable.

Peer predictors. The number of minority students per class was a significant predictor for each dependent variable except SP_LIT. This relationship was only positive for SP_LAN; for all other relationships, higher numbers of minority students predicted lower student achievement or change in achievement. Composition of students receiving free and reduced price meals was positively related to outcome literacy achievement and students' change in language performance from fall to spring, but negatively related to students' spring language achievement.

Classroom instruction and quality predictors. CLASS Instructional Support demonstrated significant relationships with all six dependent variables. This relationship was positive for all but one dependent variable (SP_LAN). ELLCO domains were more predictive of performance on curriculum-based measures than of standardized assessment performance. Lead teachers holding master's degrees helped to positively predict outcome and spring language achievement, and outcome literacy achievement. Children in classrooms with two daily small group opportunities made significantly greater gains from fall to spring on the language and literacy curriculum-based measures.

Hierarchical Linear Modeling (HLM)

Following variable reduction based on the results of the descriptive and multiple regression analyses, a two-level HLM was used to answer the research questions, which aimed to determine the relationships between student, peer and classroom instruction and quality variables on the language and literacy achievement of children in the analytic sample. HLM was used to account for the nested structure of participating children across DCPEL classrooms. This model takes into account the potential interdependence of scores for students within the same classrooms. This is important because students in the same classrooms are likely to experience the same peer, classroom instruction and quality characteristics in comparison to students in different classrooms.

At level one, student variables were entered. At level two, peer and classroom instructional quality variables were included. Six fully conditional models were estimated, one for each of the dependent variables (OUT_LAN, OUT_LIT, SP_LAN, SP_LIT, FS_LAN, FS_LIT). Because continuous variables were standardized in the preliminary data preparation stage, analysis of the slope coefficients for the models can be interpreted as standard deviation units, with a change of 1.0 indicating a 1 SD change in performance between testing periods. All predictor variables entered at levels 1 and 2 were grand mean centered.

Level 1 unconditional models. First, fully unconditional Level 1 models were employed for each of the dependent variables, with no covariates entered. The fully unconditional model is equivalent to random effects analysis of variance (ANOVA), and allowed an initial determination of significant differences between classrooms on the dependent variable entered into the model. This fully unconditional model also provides

an estimate of the intra-class correlation (ICC), which is the amount of variance that can be attributed to between class differences on the dependent variable.

Table 4.16 demonstrates the ICC across dependent variables ranged from 6.9 to 32.1 percent across dependent variables (OUT_LIT and SP_LAN, respectively). Calculating the plausible values range for OUT_LIT shows between -.44 SD to .44 SD can be attributed to between classroom differences, indicating that 95% of DCPEL classrooms fall within a 0.88 SD range in outcome language achievement. For SP_LAN, the plausible values range estimates between -1.12 to 1.10 SD, or a range of 2.2 SD in children's spring language achievement across classrooms. Thus, there is sufficient between-class variance for each model that might be estimated by adding classroom-level variables in a level 2 HLM.

Table 4.16. *Level 1 unconditional model for student language and literacy achievement in SY2009-10*

	OUT_LAN	OUT_LIT	SP_LAN	SP_LIT	FS_LAN	FS_LIT
Mean class average, γ_{00}	.037	.042	.022	.066	-0.005	.020
Between class variance	.092***	.065***	.317***	.278**	.207**	.140***
Between child variance	.825	.876	.671	.630	.811	.883
Total variance	.917	.941	.988	.908	1.002	1.023
Intra-class correlation (ICC)	.100	.069	.321	.306	.207	.147
Plausible values range	1.19 SD	.999 SD	2.21 SD	2.07 SD	1.78 SD	1.34 SD
Reliability	.618	.519	.872	.864	.787	.697

***p<.001; **p<.01; *p<.05

Level 1 conditional models. Next, conditional level 1 models were entered for each dependent variable, which included the independent variables found to be significant following the multiple regression analyses (Tables 4.10 - 4.15). The conditional level 1 model, or the within-classrooms model, tested the hypothesis that child variables affected class mean differences on the dependent variable, or if the average of variables within each classroom affected an individual student's achievement

on the dependent variable. All independent variables were grand mean centered in each hierarchical model. Table 4.17 provides results from the within-classrooms models for the six dependent variables. Predictors found significant from the previous MR analysis were grand mean centered in the HLM for each dependent variable.

Information on the plausible values range for each model, model reliability, and remaining variance is also presented in Table 4.17. Variance between classrooms remained statistically significant for all dependent variables after controlling for the student variables found to be significant during the MR analysis. This supports the use of a nested model, which accounts for the effects of students' classrooms on their language and literacy achievement.

The proportion of within-class variance explained by the student variables entered in the level 1 conditional model was estimated (Table 4.16). Student-level variables for the norm-referenced assessments accounted for 57.5 (OUT_LAN) and 58.4 percent (OUT_LIT) of the variance in language and literacy achievement within classrooms. For spring achievement on the criterion measures, student variables accounted for 46.9 (SP_LAN) and 39.4 percent (SP_LIT) of the variance within classrooms. Child variables explained a much smaller proportion of the within-classroom variance for fall-to-spring change in achievement on the criterion measures. Only 1.2 percent of the within-classroom variance in fall-to-spring language (FS_LAN) gains was attributed to child variables, and 12.6 percent of the within-classroom variance in fall-to-spring literacy (FS_LIT) gains was explained by child variables. Child-level variables were less useful in explaining within classroom differences in fall-to-spring language and literacy gains than for spring and end-of-year achievement.

Independent variables demonstrated as significantly related to the dependent variable at level 1 were retained for the final level 2 fully conditional models. All fully conditional models had sufficient level 2 units (classrooms) and students within classrooms relative to the number of explanatory variables entered into the model, so models were examined using robust standard errors. Using robust standard errors ensures that the standard errors determined by the model are correct even if HLM assumptions have been violated. Comparing the standard errors based on the model to the robust standard errors revealed no significant standard error differences. This indicates HLM assumptions were likely met.

Norm-referenced child-level predictors. Children's baseline performance on the norm-referenced language measures helped to significantly and positively predict students' spring achievement on the criterion measure and outcome achievement on the norm-referenced measures. Baseline literacy achievement on the norm-referenced assessments was a significant predictor of outcome literacy achievement on the same measures, and spring literacy achievement, and fall-to-spring change in achievement, on the criterion-referenced literacy measures. The relationship between baseline literacy performance and fall-to-spring change in performance on the progress monitoring measures was negative, indicating that a one standard deviation increase in baseline achievement on the norm-referenced assessments resulted in a -0.38 SD decrease in fall-to-spring gains on the criterion literacy assessments.

Criterion-referenced child-level predictors. Children's spring achievement on the criterion measures predicted outcome achievement on the norm-referenced measures. For language, for every one standard deviation increase in spring language achievement,

students' outcome achievement on the norm-referenced measures increased by 0.12 SD. Spring literacy achievement on the criterion measures demonstrated an even stronger effect on outcome achievement. A 0.59 SD increase in outcome performance on the norm-referenced literacy assessments was predicted by a one standard deviation increase on spring criterion measures.

Children with greater fall-to-spring change in achievement on the criterion language measure had lower outcome achievement on the norm-referenced assessments (-0.15 SD). However, fall-to-spring change in language achievement was positively associated with spring language achievement on the same criterion measure (0.45 SD). Likewise, children's fall-to-spring change in achievement on the criterion literacy assessments helped to positively predict their spring achievement on the same measures (0.29 SD).

Demographic child-level predictors. After controlling for baseline achievement on the norm-referenced measures, and where applicable spring achievement or fall-to-spring change in achievement on the criterion measures, several demographic variables were found to significantly predict students' achievement or change in achievement on the criterion measures. Controlling for entering achievement, four-year old children's standard scores on the norm-referenced outcome measures tended to be lower than their three-year old peers (OUT_LAN=-0.34 SD; OUT_LIT=-0.62 SD), while four-year old children performed better than their three-year old peers on the spring language criterion measure (0.39 SD). Female students had slightly higher outcome literacy performance than males (-0.08 SD). Students' lunch status was a significant negative predictor of outcome literacy achievement on the norm-referenced measure (FREE=-0.24 SD;

REDUCED=-0.37 SD) and spring literacy performance on the criterion measures (FREE=-0.15 SD).

ELL students performed better on the norm-referenced outcome language measures than English-only students (0.36 SD). Special education students performed lower than their typically developing peers on the spring criterion literacy assessments (-0.19 SD), but demonstrated more fall-to-spring growth (0.56 SD). Finally, students in their second program year performed less well than students in their first program year on the norm-referenced outcome language assessments (-0.22 SD).

Table 4.17. *Level 1 conditional model for student language and literacy achievement in SY2009-10*

	OUT_LAN	OUT_LIT	SP_LAN	SP_LIT	FS_LAN	FS_LIT
Fixed coefficients						
Prior achievement						
BASE_LAN	0.667***	---	0.406***	---	0.066	---
SP_LAN	0.115**	---	---	---	---	---
FS_LAN	-0.154***	---	0.451***	---	---	---
BASE_LIT	---	0.328***	---	0.536***	---	-0.379***
SP_LIT	---	0.589***	---	---	---	---
FS_LIT	---		---	0.286***	---	---
Demographic						
FOUR	-0.336***	-0.617***	0.388***			
FEMALE		0.082*				
FREE	-0.104	-0.239**	0.046	-0.152*		
REDUCED		-0.371***				
ELL	0.363***					
YR_TWO	-0.223*					
TIER3				-0.188	0.564**	
Mean class average, γ_{00}	.037	.041	.030	.066	-0.005	.016
Between class variance	.040***	.057***	.142***	.254***	.216***	.087***
Between child variance	.351	.364	.356	.382	.801	.772
Total variance	.391	.421	.498	.636	1.01	.859
Plausible values range	.784 SD	.936 SD	1.48 SD	1.98 SD	1.82 SD	1.16 SD
Reliability of random coefficients	.624	.695	.853	.905	.797	.622
Proportion of variance explained by level 1 conditional model when compared to level 1 unconditional model						
Child	57.5%	58.4%	46.9%	39.4%	1.2%	12.6%

***p≤.001; p≤.01; *p≤.05; ---independent variable never entered in MR or HLM to predict dependent variable

Level 2 fully conditional models. At Step 2, the model includes achievement on the dependent variable, controlling for pretest data, and calculated in relationship to child (Level 1) and classroom (Level 2) independent (predictor) variables. Significant coefficients estimate the magnitude of the relationship between the independent variables

and mean achievement or change in achievement of the dependent variable. Level 2 models were used to answer the four research questions explored in this dissertation.

***Research Question (RQ) 1*: To what extent are classroom peer characteristics associated with students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?** A hierarchal linear model was used to adjust for level 2 fixed effects found significant from the previous multiple regression analyses predicting outcome language and literacy achievement on the global, norm-referenced assessments. Table 4.18 demonstrates the results of the HLM allowed for rejection of the null hypothesis for RQ1 related to language and acceptance of the alternate hypothesis, that is classroom peer characteristics do affect student language outcomes. For outcome literacy, the null hypothesis was accepted. The fully conditional HLM for OUT_LIT estimated no significant relationships between classroom peer characteristics and end-of-year student literacy achievement.

Norm-referenced language (OUT_LAN). The outcome language factor consists of children's outcome scores on the norm-referenced EVT-III, PPVT-IV, and TOPEL-DV assessments. Analysis of the HLM for OUT_LAN shows that one peer variable predicted outcome language achievement for the analytic sample. The number of students who received free or reduced priced lunch per classroom was significantly and negatively related to outcome achievement in language (β=-0.11 SD). This was the only significant peer predictor from the model.

Norm-referenced literacy (OUT_LIT). No significant peer relationships were demonstrated in the level 2 HLM.

***RQ2*: What is the relationship between teacher characteristics and students' performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools?**

Examination of Table 4.18 leads to rejection of the null hypothesis and acceptance of the alternate hypothesis for RQ2. Teacher instruction and quality variables were significantly related to outcome student achievement for language and literacy.

Norm-referenced language (OUT_LAN). One general quality variable and one formal quality variable predicted student language achievement on the global outcome measures. Lead teachers' degree status was a significant predictor of outcome language achievement, with teachers with master's degrees corresponding to a 0.22 SD increase in norm-referenced language achievement. For every standard deviation increase in teachers' fall-to-spring change in CLASS Instructional Support ratings, students' outcome language scores increased by 0.16 SD. This finding seems to align with the five studies reviewed previously that found significant positive relationships between teachers' CLASS Instructional Support ratings and students' language achievement (Burchinal et al., 2008; Curby et al. 2009; Early et al., 2006; Howes et al., 2008; Mashburn et al., 2008), although previous studies did not evaluate change in ratings on the formal quality measures.

Norm-referenced literacy (OUT_LIT). One general quality variable and two formal quality variables predicted student literacy achievement on the global outcome measures. As with outcome language, lead teachers' degree status was also a significant predictor of outcome literacy achievement (0.28 SD). For every standard deviation increase in teachers' fall-to-spring change in CLASS Instructional Support ratings,

students' outcome literacy scores increased by 0.11 SD. Alternately, teachers' change in rating on the CLASS Classroom Organization domain was associated with a 0.11 decrease in outcome literacy achievement.

Table 4.18. *Level 2 fully conditional model estimating outcome language and literacy achievement for the 2009-10 school year*

OUT_LAN			OUT_LIT		
	Coefficient			Coefficient	
Fixed Coefficients	(β)	SE	**Fixed Coefficients**	(β)	SE
Intercept, γ_{00}	0.050	0.037	Intercept, γ_{01}	-0.263**	0.073
Student			*Student*		
BASE_LAN, γ_{10}	0.589***	0.039	BASE_LIT, γ_{10}	0.348***	0.043
FS_LAN, γ_{20}	-0.096**	0.038	SP_LIT, γ_{20}	0.571***	0.046
SP_LAN, γ_{30}	0.001	0.054	FOUR, γ_{30}	-0.615***	0.100
FOUR, γ_{40}	-0.055	0.112	FREE, γ_{40}	-0.255**	0.088
ELL, γ_{50}	0.362**	0.149	REDUCED, γ_{50}	-0.398***	0.092
YR_TWO, γ_{60}	-0.134	0.091			
Peer			*PEER*		
NELL, γ_{01}	0.049	0.046	NMIN, γ_{02}	-0.112	0.064
NMIN, γ_{02}	-0.113*	0.053	NFRP, γ_{03}	0.056	0.075
Classroom Instruction/Quality			*Classroom Instruction/Quality*		
LD_DGREE, γ_{03}	0.221*	0.107	LD_DGREE, γ_{04}	0.282**	0.110
FS_CL_IS, γ_{04}	0.164***	0.044	FS_CL_CO, γ_{05}	-0.111**	0.041
FS_EL_BB, γ_{05}	-0.046	0.046	FS_CL_IS, γ_{06}	0.113**	0.038
Reliability of random coefficients			**Reliability of random coefficients**		
Intercept, β_0	0.381		Intercept, β_0	0.559	
Variance Components			**Variance Components**		
Intercept, μ_0	0.023*		Intercept, μ_0	0.032***	
Level-1, r_0	0.560		Level-1, r_0	0.365	

***p≤.001; p≤.01; *p≤.05

***RQ3*: To what extent are classroom peer characteristics associated with students' spring language and literacy achievement or fall-to-spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?** RQ3 addresses relationships between classroom peer characteristics and spring language and literacy achievement on the progress monitoring measures, or change in language and literacy achievement from fall to spring. Analysis of Table 4.19 demonstrates that classroom peer characteristics significantly predicted spring language and literacy achievement and change in language

achievement from fall to spring. Thus, the null hypothesis for RQ3 is rejected for SP_LAN, SP_LIT, and FS_LAN. The alternate hypothesis, that classroom peer characteristics affects student language or literacy achievement on progress monitoring measures, or change in student achievement on language or literacy measures, is accepted. For FS_LIT, classroom peer variables did not predict change in achievement from fall to spring on the literacy factor variable, so the null hypothesis is rejected for FS_LIT.

Criterion language - spring (SP_LAN). Spring language is the standardized, continuous value (mean=0, SD=1) of children's scores on the spring administration of the IGDI-Picture Naming, a curriculum-based assessment. Three peer dependent variables significantly predicted children's spring language achievement. Number of ELL students per class predicted lower spring language achievement (-0.22 SD), but number of minority students per class helped to positively predict spring language achievement (0.24 SD). Students in classrooms with higher numbers of children who received free or reduced price lunch on average performed -0.39 SD lower than their peers in classrooms with higher numbers of children paying full price for lunch.

Criterion literacy - spring (SP_LIT). Spring literacy is a standardized, continuous factor variable predicted from students' spring achievement scores on four curriculum-based measures: PALS-LN, PALS-LS, PALS-NW, and GRTR (mean=0; SD=1). One peer variable significantly predicted 3- and 4-year-olds' spring literacy achievement. The number of returning children per class, or the number of students in their second DCPEL program year (NYR_TWO), was the only classroom peer variable that retained significance in the MR and level 1 unconditional models. Examination of the level 2

fully conditional model demonstrates that higher numbers of returning children per classroom was positively related to individual students' spring literacy achievement (β=0.24 SD).

Criterion language - fall-to-spring change (FS_LAN). Three peer variables significantly predicted change in language achievement on the IGDI-Picture Naming assessment from fall to spring. Two of the peer variables demonstrated a negative effect on students' change in achievement, while the single student variable demonstrated a positive relationship to students' change in language achievement. Children's spring language achievement was significantly related to their classrooms' average baseline language achievement. A classroom's average baseline language achievement was negatively related to a student's spring language achievement (β=-0.19 SD). This estimates that students in classrooms whose peers had higher average baseline language achievement demonstrated less fall-to-spring change on the language measure than students in classrooms with peers who began with lower average language achievement.

The number of English language learners per classroom demonstrated a positive slope at 0.21 SD. This reveals that the number of ELL students per classroom positively predicted fall-to-spring change in individual student achievement on the language measure. Students nested in classrooms with higher numbers of ELL demonstrated greater change in language achievement on the progress monitoring assessment than their peers in classrooms with lower numbers of ELL students. Finally, the number of minority students per class was negatively related to individual student change in language achievement from fall to spring (-0.22 SD). This relationship mirrors that

demonstrated on the spring language measure, with NELL and NMIN demonstrating opposite directions but similar magnitudes of relationships to the dependent variable.

Criterion literacy - fall-to-spring change (FS_LIT). No classroom peer variables were found to be significant predictors of fall-to-spring change in literacy achievement during the multiple regression analysis, so none were retained for the HLM. The null hypothesis for RQ4 related to fall-to-spring literacy change was accepted.

***RQ4*: What is the relationship between teacher characteristics and students' spring language and literacy achievement or fall-to-spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools?** Table 4.19 provides information on the fully conditional level 2 hierarchal linear models built to estimate relationships between predictor variables and student achievement, and change in achievement, on the language and literacy related criterion-referenced progress monitoring assessments. Change in quality of teacher-child interactions is significantly related to achievement on the curriculum-based language measure, and change in achievement on both the language and literacy measures. Thus, the null hypothesis for RQ4 is rejected for SP_LAN, FS_LAN, and FS_LIT. For SP_LIT, change in quality of teacher-child interactions did not predict spring achievement on the literacy factor variable, so the null hypothesis is accepted for SP_LIT.

Criterion language - spring (SP_LAN). Spring language is the standardized, continuous value (mean=0, SD=1) of children's scores on the spring administration of the IGDI-Picture Naming assessment. Children's spring language achievement was predicted by one general quality and one formal quality variable. The lead teacher's highest degree

was significantly and positively related to child spring language achievement (β=0.32 SD). Children's achievement on the spring language assessment was significantly and positively related to teachers' fall-to-spring change in rating on the ELLCO Books & Book Reading domain (AV_EL_BB; β=0.24 SD).

Criterion literacy - spring (SP_LIT). Spring literacy is a factor variable predicted from students' spring achievement scores on PALS-LN, PALS-LS, PALS-NW, and GRTR. These scores were created as a standardized, continuous variable (mean=0; SD=1). Although teachers' change in ratings from fall to spring on the CLASS Classroom Organization and Instructional Support domains demonstrated a significant relationship to spring literacy achievement through the multiple regression analysis and in the level 1 conditional model (ANOVA), these relationships were no longer significant at level 2. Thus, the null hypothesis is accepted for spring literacy achievement.

Criterion language - fall-to-spring change (FS_LAN). The fall-to-spring change variable for language is a continuous, standardized variable that subtracts students' spring achievement on IGDI-Picture Naming from their fall achievement on the same measure to create a change variable (mean=0; SD=1). Four classroom instruction and quality variables predicted students' change in language achievement from fall to spring. Change in language achievement from fall to spring was significantly related to the number of small group experiences students were provided daily. Students who were exposed to two small groups per day made 0.90 SD greater gains on the language measure between the fall and spring administrations than their peers who experienced one daily small group.

Teachers' fall-to-spring change in ratings on the ELLCO Classroom Structure and ELLCO Curriculum domains helped to positively predict students' fall-to-spring gains on the curriculum-based language measure. For every one standard deviation increase in ratings on FS_EL_CS and FS_EL_CU, students' change in achievement from fall to spring increased by 0.25 SD and 0.24 SD respectively. Students' change in achievement was negatively predicted by teachers' change in rating on the ELLCO Print & Early Writing domain. Every one standard deviation increase in rating on FS_EL_PE predicted a 0.41 SD decrease in students' fall-to-spring change in language achievement.

Criterion literacy - fall-to-spring change **(FS_LIT)**. The fall-to-spring change variable for literacy is a continuous, standardized variable that subtracts students' spring achievement on the factor variable consisting of PALS-LN, PALS-LS, PALS-NW, and GRTR from their fall achievement on the same factor variable to create a change variable (mean=0; SD=1). Children in classrooms with higher average ratings on the CLASS Instructional Support domain achieved greater gains from fall to spring on the literacy-related progress monitoring assessments, with a positive slope of 0.15 SD. This was the only significant classroom instruction and quality variable for the FS_LIT dependent variable.

Table 4.19. *Level 2 fully conditional model estimating progress monitoring language and literacy achievement and change for the 2009-10 school year*

SP_LAN			SP_LIT		
	Coefficient			Coefficient	
Fixed Coefficients	(β)	SE	**Fixed Coefficients**	(β)	SE
INTRCPT2, γ_{00}	0.031	0.074	INTRCPT2, γ_{00}	0.023	0.045
Student			*Student*		
BASE_LAN, γ_{10}	0.394***	0.312	BASE_LIT, γ_{10}	0.546***	0.042
FS_LAN, γ_{20}	0.452***	0.044	FS_LIT, γ_{20}	0.290***	0.049
FOUR, γ_{30}	0.413***	0.119	FREE, γ_{30}	-0.149*	0.066
Peer			*Peer*		
NELL, γ_{01}	-0.220**	0.072	NYR_TWO, γ_{01}	0.243***	0.052
NMIN, γ_{02}	0.240*	0.095			
NFRP, γ_{03}	-0.386**	0.106			
Classroom Instruction/Quality			*Classroom Instruction/Quality*		
LD_DGREE, γ_{04}	0.319*	0.148	FS_CL_CO, γ_{02}	-0.138	0.093
FS_CL_ES, γ_{05}	-0.054	0.083	FS_CL_IS, γ_{03}	0.134	0.088
FS_CL_IS, γ_{06}	-0.077	0.060			
FS_EL_CU, γ_{08}	-0.154	0.094			
FS_EL_BB, γ_{12}	0.239*	0.108			
Reliability of random coefficients			**Reliability of random coefficients**		
Intercept, β_0	0.821		Intercept, β_0	0.883	
Variance Components			**Variance Components**		
Intercept, μ_0	0.112**		Intercept, μ_0	0.200**	
Level-1, r_0	0.355		Level-1, r_0	0.382	

FS_LAN			FS_LIT		
	Coefficient			Coefficient	
Fixed Coefficients	(β)	SE	**Fixed Coefficients**	(β)	SE
Intercept, γ_{00}	0.011	0.065	Intercept, γ_{00}	0.015	0.059
Student			*Student*		
TIER3, γ_{10}	0.510**	0.199	BASE_LIT, γ_{10}	-0.371***	0.046
Peer			*Peer*		
NELL, γ_{01}	0.209*	0.099			
NMIN, γ_{02}	-0.222*	0.103			
NFRP, γ_{03}	-0.147	0.127			
AVGLAN, γ_{04}	-0.192**	0.081			
Classroom Instruction/Quality			*Classroom Instruction/Quality*		
SM_GRP, γ_{05}	0.896***	0.231	SM_GRP, γ_{02}	0.169	0.147
FS_CL_IS, γ_{06}	0.141	0.081	FS_CL_IS, γ_{03}	0.147**	0.050
FS_EL_CS, γ_{07}	0.252*	0.115			
FS_EL_CU, γ_{08}	0.242*	0.092			
FS_EL_BB, γ_{09}	-0.178	0.110			
FS_EL_PE, γ_{10}	-0.408***	0.093			
Reliability of random coefficients			**Reliability of random coefficients**		
Intercept, β_0	0.724		Intercept, β_0	0.546	
Variance Components			**Variance Components**		
Intercept, μ_0	0.145***		Intercept, μ_0	0.064***	
Level-1, r_0	0.802		Level-1, r_0	0.771	

***p≤.001; p≤.01; *p≤.05

Comparison of models at levels 1 and 2. Table 4.20 presents proportion of variance estimates for the level 2 HLM. Examination of the variance components at levels 1 and 2 for the six dependent variables demonstrates that the peer and classroom instruction and quality independent variables introduced at level 2 reduce the proportion of between classroom variance between 21.1 (SP_LAN) and 43.9 percent (OUT_LIT).

Table 4.20. *Estimates for proportion of variance explained by level 2 hierarchical linear models for the dependent variables*

	OUT_LAN	OUT_LIT	SP_LAN	SP_LIT	FS_LAN	FS_LIT
Level 1 between class variance	.040***	.057***	.142***	.254***	.214***	.087***
Level 2 between class variance	.023**	.032**	.112**	.200**	.145***	.064***
Proportion of variance explained by Level 2 model	42.5%	43.9%	21.1%	21.3%	32.2%	26.4%

Summary of Analyses

Results from principal components analysis, descriptive statistics, multiple regression analysis, and 2-level hierarchical linear models were used to answer the four research questions.

Principal components analysis (PCA). In order to reduce the number of dependent variables used in this research, PCA was implemented to create factor variables according to the language and literacy constructs determined a priori using data from the established psychometrically sound norm- and criterion-referenced assessments administered to students in the 2009-10 school year. All factors loaded with adequate coefficient alphas.

Descriptive statistics. Results from the descriptive analyses indicated small to large effect sizes on student achievement for three and four-year old children on both norm-referenced and criterion-referenced assessments. Likewise, the quality of teacher-child interactions across classrooms demonstrated mostly large effect sizes ($d \geq 0.80$).

However, there were significant differences in formative and summative achievement and change in achievement on the norm-referenced and criterion-referenced assessments between children with free and reduced price lunch status and their paying peers. Only spring language and fall-to-spring change in language on the criterion measure did not demonstrate differences based on lunch status.

Significant variance was demonstrated between the quality of teacher-child interactions experienced by children whose families paid full price for lunch and their peers who received free or reduced price lunch. For six of the eight CLASS and ELLCO domains, children whose families paid full price for lunch were in classrooms with teachers rated higher on the teacher-child interaction measures. Children received similar exposure to emotional support and classroom organization regardless of lunch status, as measured by the CLASS.

Multiple regression (MR). Multiple regression analyses using a generous $p \leq .15$ supported a reduction of predictor variables for each of the dependent variables. Student, peer and classroom instruction and quality variables that remained significant from the MR analyses were entered into the subsequent hierarchical linear models used to answer the research questions.

Hierarchical linear modeling (HLM). The first step in the HLM analysis involved estimating the intraclass correlation coefficient (ICC) for each dependent variable to determine if sufficient variance existed across classrooms to warrant the use of nested data models. The ICC ranged from 6.9 to 32.1 percent across dependent variables (OUT_LIT and SP_LAN, respectively), which was sufficient between class variance to support further analysis using HLM.

Level 1 conditional models were then created to estimate the within classroom variance on each of the dependent variables that could be attributed to the child-level predictors entered into each model. Student-level variables used to predict the norm- and criterion-referenced assessments accounted for a significant portion of the variance in language and literacy achievement within classrooms. Student-level variables explained a much smaller proportion of the within-classroom variance for fall-to-spring change in achievement on the criterion measures.

Using the information from Step 1, Step 2 involved entering a fully conditional level 2 HLM for each dependent variable. These models were used to estimate the effect peer and classroom instruction and quality characteristics had on children's performance on the dependent variables after controlling for the level 1 student variables.

Peer variables. Five of the nine peer variables demonstrated significant relationships with the dependent variables analyzed in this study. The number of four-year-old children, number of females, average baseline literacy achievement, and number of special education students were not significant predictors of student achievement, or change in achievement, for any of the dependent variables. The number of minority students per classroom was positively related to students' spring language achievement, but negatively related to students' outcome language achievement on the standardized measures and change in language achievement from fall to spring on the curriculum-based assessment.

Average baseline language achievement was negatively related to fall-to-spring change in language achievement. The number of ELL students and number of students receiving free and reduced price lunch per classroom was negatively related to children's

spring language achievement on the IGDI-Picture Naming assessment. The number of second year students was positively related to spring literacy.

Classroom instruction and quality variables. Eight of the ten classroom instruction and quality variables demonstrated significant relationships with the dependent variables used in this study.

Classroom instruction variables. The variable created to estimate the quantity of interaction an individual child might receive with a classroom teacher was found to be significant - number of daily small groups. The results of this research indicate that children in classrooms that provided two daily small group learning opportunities made greater gains from fall to spring on the curriculum-based language assessment.

Classroom quality variables. On average, children in classrooms where lead teachers had master's degrees outperformed their peers from classrooms with one daily small group in norm-referenced language, norm-referenced literacy, and in criterion language. With the exception of the CLASS Emotional Support and ELLCO Language Environment domains, all of the variables entered to estimate change in quality of teacher-child interactions were found to significantly predict student achievement in language and literacy.

CLASS domain predictors. Children whose teachers demonstrated greater change on CLASS Classroom Organization (FS_CL_CO) ratings demonstrated lower norm-referenced literacy achievement than their peers in classrooms with teachers demonstrating less change. Children's norm-referenced language and literacy achievement was positively related to teachers' change in ratings on the CLASS Instructional Support domain. Surprisingly, based on the previously reviewed studies,

children whose teachers had higher gains in CLASS Instructional Support ratings demonstrated lower outcome language achievement than their peers in classrooms with teachers making fewer gains.

ELLCO domain predictors. Teachers' change in ratings on three of the four significant ELLCO domains helped to positively predict student achievement and change in achievement on the curriculum-based measures. ELLCO domains did not help to predict standardized achievement in language or literacy. A positive change in teachers' ratings on the ELLCO Classroom Structure and Curriculum domains predicted gains in children's fall-to-spring criterion language achievement when compared with children in classrooms whose teachers' ratings remained constant. Children in classrooms whose teachers' demonstrated more fall-to-spring change in rating on the ELLCO Books & Book Reading domains outperformed their peers on the spring criterion language assessment when compared with children from classrooms where teachers demonstrated less change in this domain. Interestingly, teachers' change in rating on the ELLCO Print & Early Writing domain demonstrated a significant but negative relationship with students' change in criterion literacy achievement from fall to spring.

Conclusions. Based on the statistical analyses for data from the analytic sample, the null hypothesis was rejected for norm-referenced outcome language, criterion spring language, and criterion fall-to-spring change in language achievement for each research question. Peer and classroom instruction and quality variables helped to explain between classroom variance for students' scores on OUT_LAN, SP_LAN and FS_LAN. Norm-referenced literacy achievement (RQ1) and change in criterion literacy achievement from fall to spring on the progress monitoring assessments (RQ3) were not significantly related

to any peer variables based on the level 2 HLM. Children's spring literacy achievement on the criterion assessments was not related to classroom instruction and quality characteristics. Table 4.21 provides a visual summary of the significant variables across the six 2-level HLM models.

Table 4.21. *Coefficients for predictor variables found significant across the study dependent variables*

	Variable Name	OUT_LAN	OUT_LIT	SP_LAN	SP_LIT	FS_LAN	FS_LIT
Dependent Variables							
Language	SP_LAN						
	FS_LAN	-0.096**		0.452***			
Literacy	SP_LIT		0.571***				
	FS_LIT				0.290***		
Independent Variables							
Student - Academic	BASE_LAN	0.589***		0.394***			
	BASE_LIT		0.348***		0.546***		-0.371***
	ELL	0.362**					
	TIER3					0.510**	
	YR_TWO						
Student - Demographic	FREE		-0.255**		-0.149*		
	REDUCED		-0.398***				
	RACE						
	FEMALE						
	FOUR		-0.615***	0.413***			
Peer	AVGLIT						
	AVGLAN					-0.192**	
	NELL			-0.220**		0.209*	
	NTIER3						
	NYR_TWO				0.243***		
	NFOUR						
	NFEMALE						
	NMIN	-0.113*		0.240*		-0.222*	
	NFRP			-0.386**			
Classroom instruction and quality	LD_DGREE	0.221*	0.282**	0.319*			
	SM_GRP					0.896***	
	FS_CL_ES						
	FS_CL_CO		-0.111**				
	FS_CL_IS	0.164***	0.113**				0.147**
	FS_EL_CS					0.252*	
	FS_EL_CU					0.242*	
	FS_EL_LE						
	FS_EL_BB			0.239*			
	FS_EL_PE					-0.408***	

***p≤.001; p≤.01; *p≤.05

Chapter 5: Discussion & Conclusions

The aim of this dissertation research was to expand on current research that links student achievement, as measured by norm- or criterion-referenced assessments, to classroom-level factors including peer characteristics and quality of teacher-child interactions for preschool children in a variety of publicly funded early childhood education settings (Curby et al., 2009; Early et al., 2007; Gallagher & Lambert, 2006; Howes et al., 2008; Mashburn et al., 2008). Specifically, the research questions for this study built on previous research by exploring whether peer variables, such as the number of special education and ELL students per classroom, and classroom characteristics, such as teacher education and change from fall to spring in ratings of teachers' instructional quality, helped to explain differences in the achievement of both three and four-year old children attending public charter schools on formative and summative language and literacy assessments. Previous research analyzed teachers' quality at one point in time, which might not accurately approximate the quality children experienced throughout the year. Previous research was also conducted in traditional public schools, Head Start classrooms, and publicly funded classrooms in community-based organizations, so no previous studies analyzed these relationships in public charter school programs. Two studies included special education and or ELL students, but both used only criterion-referenced assessments as the dependent variables. The design of the current study addressed these issues.

The decision to use change variables for the teacher quality measures rather than point-in-time variables is one possible explanation that some findings from this research do not align with findings from previous studies. All studies reviewed for this

dissertation used point-in-time estimates of quality on the CLASS to analyze relationships with student achievement (Burchinal et al., 2009; Curby et al., 2009; Early et al., 2006; Early et al., 2007; Howes et al., 2008; Mashburn et al., 2008). While change variables introduce greater potential for measurement error, good reliability of the measures used is important to limit possible measurement error. Teacher quality measures used in this study demonstrated strong reliability. In addition, assessors trained to reliability using the methods established by the tools' researchers collected all teacher quality data used in this study. Knowledge of how teachers' change in performance over time relates to students' language and literacy achievement is valuable to more closely inform how students' experience of quality over the course of the school year impacted their achievement (Burchinal et al., 2008). Average ratings may mask the magnitude of differences between children's experiences at different points in time, and measurements of change can potentially be used to estimate effectiveness of professional development provided to teachers (Neuman & Cunningham, 2009), which was a major component of the Early Reading First project that funded data for this study.

Another design decision that may have affected differences in outcomes between the current study and previous research is to include multiple dependent variables. However, the use of multiple measures to assess achievement in a construct is a more valid method than use of a single measure. Furthermore, data from this research met the necessary assumptions for the methods used in this study. All student assessments and teacher observations had strong psychometric properties, and the factor loadings from the principal components analysis used to reduce the set of dependent variables were quite strong.

Classroom peer characteristics used as predictor variables in this study included the average initial language and literacy achievement of the class on the norm-referenced measures, the number of ELL or special education students per class, the number of students in their second program year, the number of four-year old or female students, the number of minority students, and the number of students receiving free or reduced price lunch. Classroom instruction and quality variables were lead teacher's degree level, number of daily small group instructional experiences, teachers' fall-to-spring change in performance on the CLASS Emotional Support, Classroom Organization, and Instructional Support domains, and teachers' fall-to-spring change in performance on the ELLCO Classroom Structure, Curriculum, Language Environment, Books & Book Reading, and Print & Early Writing domains.

The results of this dissertation estimated classroom-level effects that are generally consistent with the identified conceptual framework and outcomes of the previous studies reviewed to inform this research. Building on seminal ecological systems theory work by Brofenbrenner and Morris (1988), the conceptual framework for this study posited that classroom and teacher characteristics act as moderators for individual students' language and literacy achievement (Justice & Sofka, 2010). Outcomes from previous research with four-year olds have demonstrated that peer and classroom instruction and quality characteristics were differentially predictive of students' language and literacy achievement on norm- and criterion-referenced assessments (Curby et al., 2009; Early et al., 2007; Gallagher & Lambert, 2006; Howes et al., 2008; Mashburn et al., 2008). Findings from the current study demonstrate that teachers' change in quality as measured by two validated observational tools, which might be malleable due to influences such as

professional development (Neuman, 2009; Neuman & Cunningham, 2009), can significantly impact students' language and literacy achievement as demonstrated by norm- and criterion-referenced measures.

First, answers to the research questions are discussed in relationship to peer and classroom instruction and quality characteristics, and linked to previous findings from the literature reviewed to inform this study. Next, potential implications are presented. The final pages provide study limitations and conclusions.

Findings from the Current Study

This study analyzed how peer characteristics and change in teacher quality related to students' achievement, or change in achievement, on norm- and criterion-referenced language and literacy measures. Performance on both assessments was analyzed because of the differences in purpose for the two types of assessments, and because this research sought to analyze how teacher quality and peer characteristics impacted student achievement. Norm-referenced assessments are created to assess relative performance on a large construct as a far transfer measure; the assessments are administered infrequently and do not necessarily reflect actual skills and concepts taught in the child's classroom. Criterion-referenced assessments, such as the curriculum-based measures used in this research, offer a viable method to systematically and frequently assess student strengths and weaknesses and change in performance (Horowitz, 2006), by their nature of being brief, repeated assessment measures focused on specific domains or skills. In addition, criterion assessments are created to be sensitive to the effects of instruction. Understanding if teacher or peer characteristics are significant predictors of student

performance on the different types of assessments provides information about different types of knowledge supported by classroom variables.

Norm-referenced assessments. RQ1 and RQ2 explored the relationships between peer and classroom instruction and teacher quality characteristics and three and four-year old children's end-of-year (outcome) performance on norm-referenced language and literacy assessments.

Research Question (RQ) 1*: *To what extent are classroom peer characteristics associated with student performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools? A significant relationship was demonstrated between peer characteristics and students' performance on the norm-referenced language factor variable, which consisted of two expressive and one receptive language assessment. Children in classrooms with lower numbers of minority students outperformed their peers in classrooms with higher numbers of minority students on both outcome measures. No other peer characteristics were found to be significant predictors of student language achievement. For the norm-referenced literacy assessments, no peer characteristic significantly predicted students' outcome literacy achievement.

Previous research demonstrated that classrooms with more poor and minority students demonstrate lower student language and literacy achievement (Dotterer et al., 2009). This aligns with the findings from the current research, although no literacy relationship was demonstrated in the current analysis. Gallagher and Lambert (2006) demonstrated that lower initial peer averages on criterion measures predicts student language and literacy achievement on the same measures. The relationship between

baseline classroom averages on the norm-referenced measures and students' year-end achievement was not significant for the norm-referenced language or literacy factors created for use in this study. This difference may be because Gallagher and Lambert used criterion measures.

Research Question (RQ) 2*: *What is the relationship between teacher characteristics and student performance on standardized, norm-referenced early language and literacy outcome assessments for three and four-year olds in public charter schools? Lead teachers' degree status and teachers' change in performance on two domains of the CLASS tool helped to significantly predict students' achievement on the norm-referenced outcome language and literacy measures. Lead teachers' degree status was associated with a 0.22 SD higher outcome for children in language, and a 0.28 SD higher outcome for the norm-referenced literacy factor. Additionally, higher change in rating from fall to spring on the CLASS Instructional Support domain predicted a 0.16 SD positive difference in outcome literacy achievement. The finding regarding teacher degree status is in contrast to prior research reviewed for this study, which demonstrated that teachers' highest degree shows little relationship to student outcomes when entered with more specific quality measures such as the CLASS (Brown et al., 2008; Early et al., 2006; Early et al., 2007).

One possibility for the difference is that previous studies have not examined change in classroom quality ratings over time. Lead teachers' degree status is determined based on teachers' status at the beginning of the school year, and remained constant over the course of the school year. So, any benefits children might receive from exposure to better-educated teachers would occur throughout the entire school year. Alternately,

subtracting fall ratings on the observation tools from spring ratings derived the CLASS and ELLCO change scores. Post-hoc regression analyses for the CLASS domains *(F*(25.41) = .000 to F(123.88) = .000 and ELLCO domains *(F*(37.81) = .000 to *F*(155.80) = .000) demonstrates that teachers who had more growth on the observational quality measures began lower than their colleagues who made less growth. Thus, children in these classrooms began the year experiencing lower quality interactions than peers in classrooms with teachers who demonstrated less growth. This is one explanation for the lower magnitude of relationship between norm-referenced language and literacy achievement and change in CLASS Instructional Support ratings compared to lead teachers' degree status. Children benefited throughout the year from teachers' education status, but had differential exposure to higher quality Instructional Support as measured by the CLASS.

CLASS Classroom Organization was also a significant predictor of students' achievement on norm-referenced literacy. Children who were in classrooms with higher ratings gains from fall to spring in the Classroom Organization domain performed less well on the norm-referenced literacy measure than their peers in classrooms where teachers experienced less fall-to-spring growth in this domain. Review of teachers' performance on the CLASS Classroom Organization domain demonstrates a range of 3.45 points between the lowest and highest rated teachers in the fall, and a range of 2.79 points between the lowest and highest rated teachers in the spring. Teachers at the lower end of ratings in the fall made gains, but they maintained nearly a 3-point difference in ratings on the CO domain, which is highly significant statistically. Because teachers began with low Classroom Organization in the fall and remained low in the spring,

students in these classrooms were not consistently exposed to high quality Classroom Organization. It is possible that students who began in classrooms with lower levels of Productivity, Behavior Management, and Instructional Learning Formats, the dimensions measured in the CLASS Classroom Organization domain, missed crucial time for developing the literacy skills needed to translate to more global achievement on the norm-referenced outcome assessments.

Change in performance on the ELLCO domains was not significantly related to child language or literacy achievement on the norm-referenced measures. The number of daily small group experiences children received was also not useful in predicting students' outcome language or literacy achievement.

Summary. In combining the results of RQ1 and RQ2, classroom instruction and change in quality characteristics were more predictive of student achievement on norm-referenced language and literacy assessments than classroom peer characteristics. Early and colleagues (2007) demonstrated that quality of teacher-child interactions could predict effects on students' language and literacy development even after controlling for classroom peer characteristics. In the current study, peer characteristics were associated with 0.11 SD decrease in outcome language achievement, while classroom instruction and quality change characteristics combined to predict 0.39 SD increase in norm-referenced language achievement. For the literacy measures, no peer characteristics were found predictive of students' norm-referenced literacy achievement. Thus, this research aligns with the previous findings of Early et al. (2007), who demonstrated that teacher quality variables are stronger predictors of students' norm-referenced language and literacy achievement than are peer characteristics. This research extends those

findings to data that includes three-year old children, children in charter schools, controls for ELL and special education peer characteristics, and measures change in teacher quality over time, as well.

Criterion-referenced assessments. RQ3 and RQ4 explored the relationships between peer and classroom instruction and teacher quality characteristics and three and four-year old children's progress monitoring (spring and fall-to-spring change) performance on criterion-referenced language and literacy assessments.

Research Question (RQ) 3: To what extent are classroom peer characteristics associated with students' spring language and literacy achievement or fall-to-spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools? Five peer characteristic variables demonstrated significant relationships to the criterion language and literacy assessments analyzed in this study: average baseline language; number of ELL students per class; number of special education students per class; number of minority students per class; and number of second-year students per class. In all this accounted for seven significant relationships, of which four were negative.

Classroom average baseline achievement. Students in classrooms where peers had higher average baseline language ability demonstrated less fall-to-spring change on the curriculum-based language assessment (-0.19 SD). One explanation for this might be regression to the mean, where children with higher baseline language achievement demonstrate less language growth over the school year. Another possible explanation relates to the differences between the norm-referenced assessments administered at baseline and the criterion measure used for progress monitoring. The language

assessment administered for progress monitoring uses common pictures to assess children's language development (e.g., *fish*, *rainbow*, *whistle*), so perhaps children in classrooms where peers have higher average initial language achievement on the norm-referenced measures are exposed to more sophisticated words during regularly scheduled peer-to-peer interactions such as center time or recess.

Teachers might also use more sophisticated language in classrooms where students begin with higher average language ability. Dotterer and colleagues (2009) demonstrated that teachers in classrooms with fewer poor students rate higher on the Instructional Support domain, which focuses on the quality of language interactions between teachers and students. Data from Hart and Risley (1985) provides evidence that children from poor families are more likely to have limited vocabulary at school entry. Results from the descriptive analyses conducted in this study demonstrate mean differences in baseline language achievement based on lunch status, with students who paid full price for lunch significantly outperforming students who received free or reduced price lunch at school entry. Likewise, analysis of DCPEL data showed that children who paid full price for lunch were more likely to be in classrooms rated as higher quality on the CLASS Instructional Support domain. Thus, it is possible that children in classrooms where fewer peers received free or reduced price lunch had less exposure by students or teachers to the kind of vocabulary represented in the criterion measure.

Numbers of student subgroups per classroom. Results from this study indicate that students in classrooms with higher numbers of minority students made less fall-to-spring change on the criterion-referenced language measure (-0.22 SD) than children in

classrooms with fewer numbers of minority students. However, children in classrooms with higher numbers of minority students outperformed their peers in classrooms with fewer minority students on the spring criterion language assessment (0.24 SD). The original t-test values from the analytic sample data used in this study demonstrated that spring language achievement and fall-to-spring change in language scores were not different based on lunch status. Post-hoc analysis of t-test differences for African American students demonstrates that African American students significantly outperformed their White, Hispanic and other race peers on the fall administration of the criterion language assessment ($t=-1.93$, $p=0.02$). African American students made up the majority of the analytic sample, and were the highest subgroup of minority students in every classroom. Because minority students had higher fall scores on the criterion language assessment, regression to the mean is a reasonable explanation for this relationship.

The number of ELL students per classroom predicted a -0.22 SD change in performance on the spring criterion language measure, while the number of special education students per classroom predicted a -0.12 SD change. These findings also align with Gallagher and Lambert (2006), although the researchers used ratio rather than total number of peers with target characteristics. The researchers demonstrated that children in classrooms where teachers had high quality, but which also had higher ratios of special education students, performed less well on the assessment measures than their peers in classrooms with lower ratios of special education students. Gallagher and Lambert (2006) posited that higher numbers of students needing additional and specialized support constrains the teacher resources available to all students, and may negatively impact all

children. Another potential explanation is that the children in these classrooms began with lower language ability and despite gains did not "close the gap" on the criterion measure.

The number of four-year old students per classroom helped predict student achievement on three dependent variables, and change in fall to spring achievement on the criterion language measure. Children in classrooms with higher numbers of four-year olds performed less well on outcome achievement measures than their peers in classrooms with higher numbers of three-year old children. Specifically, number of four-year old children per class predicted -0.26 SD lower outcome language achievement, and -0.49 SD lower outcome literacy achievement.

Analysis of the relationship between number of four-year olds per class and spring literacy achievement demonstrates that number of four-year olds per classroom is positively related to students' spring criterion literacy achievement (0.39 SD). This finding is opposite of the negative relationship between number of four-year olds per class and performance on the outcome literacy measures. One explanation for this difference is that the spring literacy assessments measures more discrete skills traditionally taught in early childhood programs and in homes (e.g., letter identification, letter-sound identification, and name writing) and are criterion-referenced. The outcome literacy assessments measure more global skills (e.g., quality of print, blending, elision, and syllabication), which may not be directly taught and are norm-referenced, so the score differences may be explained by regression to the mean. Perhaps being in classrooms with larger numbers of four-year old children allows for more opportunities for peer modeling of the discrete skills assessed on the progress monitoring measures,

since four-year old students begin the year performing higher on the progress monitoring assessments than their three-year old peers.

Finally, children in classrooms with higher numbers of four-year old children demonstrated less change from fall to spring on the language measure (-0.18 SD). The criterion-based measure is a one-minute task in which children have to name as many pictures as possible after being shown flashcards by their teacher. The differences in scores might be due to a ceiling effect, because all children in classes with higher number of four year olds began with the ability to name more pictures on the picture-naming task from the fall administration, and were limited in the gains possible on a timed task.

Research Question (RQ)4: What is the relationship between teachers' change in quality and students' spring language and literacy achievement or fall-to-spring change in achievement on standardized, criterion-referenced progress monitoring assessments for three and four-year olds in public charter schools? Classroom instruction and quality variables predicted spring language, and fall-to-spring change in language and literacy on the criterion measures. Fall-to-spring change in language was predicted by four different classroom instruction and quality variables.

Lead teachers in DCPEL had either master's degrees or bachelor's degrees. Analysis of the variance explained by the lead teacher's degree demonstrated a significant relationship to only one of the criterion dependent variables. Children's spring language achievement was positively related to lead teachers' degree status (0.31 SD). This finding is in opposition to the research reviewed for this study, which demonstrated that when teacher degree status was a significant predictor of child achievement, the relationship was only demonstrated for discrete literacy skills such as letter naming (Brown et al.,

2008; Early et al., 2006; Early et al., 2007). The nature of the one-minute timed test for language progress monitoring, and the key vocabulary used in the assessment, make this vocabulary measure more discrete than the norm-referenced measures used in the reviewed studies. This may help to explain why lead teachers' degree status was significantly related to vocabulary achievement in this study. Specifically, it is possible that because the criterion language assessment is a near transfer task that uses words commonly taught and used with three and four-year old children and requires one-minute administration, this assessment may act more as a discrete test of knowledge similar to a letter identification or color naming task than the norm-referenced assessments that are designed to use more difficult vocabulary and are untimed.

Number of daily small groups demonstrated a positive, significant and large relationship to children's fall-to-spring gains on the criterion language measure. Students who received two daily small groups made 0.90 SD greater gains on the picture-naming task than their peers in classrooms that had one daily small group. Because this variable approximates a school effect, in that three Partnership schools implemented two small daily small groups and three schools implemented one small group, exploration of potential interactions with other classroom-level variables, such as average baseline language and literacy achievement, would help to explain if this relationship persists regardless of classroom-level differences. Prior research demonstrates that small group instruction provides substantially more opportunities for students to use language to ask and answer questions with teachers and peers than does large group instruction (Fletcher, 2004), and specifically benefits typically developing preschoolers and those with

language delays (Lowenthal, 1980), so it is reasonable that more exposure to small groups would support vocabulary development.

Children whose teachers demonstrated more fall-to-spring gains on the ELLCO Books & Book Reading domain had higher spring criterion language achievement than children in classrooms where teachers demonstrated less change. Teachers' change in achievement on ELLCO Classroom Structure and Curriculum domains helped predict students' change in achievement from fall to spring on the criterion language measure. A moderate negative relationship was demonstrated between teachers' change in rating from fall to spring on the ELLCO Print & Early Writing domain and students' change in achievement from fall to spring on the criterion language measure. Teachers who demonstrated greater gains in CLASS Instructional Quality were positively associated with children's fall-to-spring gains on the criterion literacy assessments.

The significant and moderate negative relationship demonstrated between teachers' fall-to-spring improvement in ratings on the ELLCO Print & Early Writing domain and students' fall-to-spring change in language achievement is interesting, with potential implications for balance of time spent on language and literacy skills. It is possible that teachers in classrooms that rated lower on ELLCO_PE from the beginning of the year spent more time focusing on building children's print awareness and providing more rich writing opportunities and did not balance this focus with strong support for language and vocabulary development. Analysis of the correlations among instructional quality change variables demonstrates lower correlations between ELLCO Print & Early Writing and the two ELLCO language-specific domains, Language Environment and Books & Book Reading Opportunities. Meaning, teachers' change in quality was more

related between the two language-related domains than with the Print & Early Writing domain. This may be associated with the in-class professional development structure employed by DCPEL. Teachers are supported in the creation of action plans to address their areas of relative weakness on CLASS and ELLCO. One possibility might be to embed maintenance strategies for domains that are not the target of action plans to ensure that teachers maintain overall quality in the pursuit of improved quality on specific areas.

Summary of classroom peer relationships to student achievement. Only one classroom peer variable predicted three and four-year olds achievement on the norm-referenced assessments. The number of minority students per classroom was negatively related to individual students' outcome scores on the norm-referenced language measures. The number of minority students was related to less change in achievement from fall to spring on the criterion language assessment, but predicted higher levels of achievement in spring on the criterion language measure.

Only one classroom peer variable predicted students' achievement related to literacy. Children in classrooms with higher numbers of children in their second DCPEL year outperformed their peers from classrooms with fewer numbers of returning students on the spring criterion literacy assessments. It is interesting that peer variables were more frequently predictive of language-related performance (7 relationships) than literacy performance (1 relationship). Perhaps teachers provide the majority of literacy inputs in the early childhood classrooms in this study, whereas teachers and students provide ongoing language exposure throughout the day. Examination of the relationships between classroom peer characteristics and student achievement on the norm- and criterion-referenced language assessments demonstrate that five of the seven relationships were

negative. Further examination of these relationships using percentages of student subgroups per classroom might provide more clarity on these relationships (Gallagher & Lambert, 2006).

Summary of classroom instruction and quality relationships to student achievement. Lead teachers' highest degree helped to positively predict achievement on three of the dependent variables. Children whose teachers had master's degrees outperformed their peers whose teachers had bachelor degrees on the norm-referenced language and literacy measures, and on the criterion language measure. Additionally, children who received two daily small group opportunities made greater gains from fall to spring on the criterion language assessment.

Summary of analyses using CLASS and ELLCO change variables. Change in performance on two formal classroom quality measures was used to estimate effects on student achievement. The CLASS tool is a measure of global teacher quality focused on general teacher-child interactions that might occur across a variety of activities during an instructional day. The ELLCO is targeted to evaluate the quality of teacher-child interactions related to language and literacy. Therefore, the researcher posited that teachers' change in performance on the ELLCO tool would explain more variance in student language and literacy achievement than would teachers' change in performance on the CLASS. For the CLASS, teachers' change in performance on the Instructional Support and Classroom Organization domains was associated with student achievement or change in achievement. For the ELLCO, teachers' change in performance on the Classroom Structure, Curriculum, Books & Book Reading Opportunities, and Print &

Early Writing domains was significantly related to students' achievement or change in achievement.

Analysis of the slope coefficients for the domains of each tool demonstrated that change in ratings on ELLCO domains was a stronger predictor of student achievement, and change in achievement, on the curriculum-based measures than was change in ratings on the CLASS domains. On the global standardized language and literacy measures, CLASS Classroom Organization and Instructional Support predicted outcome achievement and no ELLCO domains maintained significance in the final analyses. This makes sense in the context of much of the research reviewed for this study, which demonstrated that teacher-child interaction measures that are targeted to language and literacy interactions are more predictive of students' language and literacy achievement than general teacher-child interaction measures (Burchinal et al., 2008; Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008).

However, the findings from this study are more nuanced than the original supposition that the ELLCO would explain more variance in student performance than the CLASS due to the ELLCO observation's explicit focus on language and literacy instruction. Change in performance on four of the five ELLCO domains was more useful for predicting relationships with criterion measures and not norm-referenced measures, consideration of the focus for each teacher quality measure in relationship to the focus of the student assessment measures provides a potential explanation. CLASS domains analyze global teacher-child interactions that might occur during any portion of a school day, including lunch and recess. Alternately, ELLCO domains analyze language and literacy teacher-child *instructional* interactions, with specific instructions to raters to base

ratings on typical learning activities (e.g., centers, storybook reading, and writing instruction). Thus, that CLASS variables were more often significant predictors of global student outcome measures and the ELLCO variables more often significantly predicted curriculum-based measures, which were developed to be more sensitive to instructional interventions than global measures, is logical. ELLCO domains provide more information specific to teachers' language and literacy instruction, and curriculum-based assessments are created to evaluate students' mastery of items that are directly instructed.

Furthermore, teachers' change in ratings on the ELLCO domains predicted progress monitoring achievement and change in achievement on the criterion language measure. Because students' performance and change in performance on the language progress monitoring measure helps to predict achievement on the norm-referenced outcome composite, it is reasonable to posit that teachers' change in ratings on the significant ELLCO domains may help to moderate students' outcome language achievement by directly impacting the achievement on the curriculum-based measure. More research investigating possible interaction effects for these variables is warranted.

Relationships found in this study aligned with some results from the previous research and not with others. The Classroom Organization domain of the CLASS, and the Print & Early Writing domain of the ELLCO, demonstrated negative relationships with student achievement or change in achievement. Students whose teachers demonstrated greater gains in Classroom Organization demonstrated lower norm-referenced language outcome achievement. Students whose teachers made more gains in Print & Early Writing ratings demonstrated less change in criterion language achievement from fall to spring. There is no published literature on the relationship

between ELLCO domains and student achievement, and previous research using the CLASS did not include the Classroom Organization domain (Burchinal et al., 2008; Curby et al., 2009; Howes et al., 2008; Mashburn et al., 2008).

The CLASS Classroom Organization domain is a newer domain that was not used in any of the national program evaluations or secondary research reviewed here, so comparisons between previous research and the findings in this study are not possible for that domain. However, because this domain measures behavior management, productivity, and instructional learning formats, it is possible that students who begin the year exposed to lower quality behavior management, and in classrooms that are less productive and have less diverse systems for content delivery are not able to make sufficient progress later in the year as their teachers' quality improves. Connor and colleagues have demonstrated with first graders that quality and types of instruction children receive from the beginning of the year help to predict their year-end achievement (Connor et al., 2009), and it is probable that teachers who demonstrated more fall-to-spring change began lower in their fall ratings. Data from this study also suggest that teachers who demonstrated lower quality at the beginning of the year narrowed but did not close the quality gap by spring.

Summary of findings. This dissertation research demonstrated that classroom peer characteristics, instruction, teachers' background, and teachers' change in instructional quality characteristics impacted students' achievement and change in achievement on norm- and criterion-referenced language and literacy measures, after controlling for risk factors based on students' demographic backgrounds and key prior achievement characteristics.

Peer characteristics. Only one peer characteristic significantly predicted norm-referenced achievement. The number of minority students per classroom was negatively associated with children's norm-referenced language achievement. Criterion achievement in language was predicted by three peer characteristics: number of ELL, minority and free or reduced price lunch students per classroom. Change in language achievement on the criterion measure was also predicted by three peer characteristics: number of ELL and minority students per class, and the classroom average baseline language ability on the norm-referenced measure. Results indicate that children who began lower on the criterion measure made greater fall-to-spring progress in language than their peers with higher scores, but did not make sufficient gains to "close the achievement gap" by spring administration of the measure. Only one peer characteristic helped to predict criterion literacy achievement. The number of students in their second program year per class was positively related to criterion literacy achievement.

Teacher characteristics. The teacher variables most related to student outcomes were change in quality of teacher-child interactions, and teachers' highest degree. This finding is contradictory to previous research that did not demonstrate strong relationships between teachers' highest degree and student achievement, after controlling for teacher quality(Brown et al, 2008; Early et al., 2006; Early et al., 2007) . One possibility for this difference is that data from the current study comes from a program that provides a standard language and literacy curriculum to all classrooms. Additionally, teachers in this project received the same training, and ongoing professional development from a single cadre of classroom coaches. Findings from the literature review indicate that curriculum and professional development were not controlled for in the majority of

studies. These program-level inputs may have had an impact on the relationship between teachers' degree status and students' language and literacy achievement.

Analysis of data from this study also demonstrated that teachers' change in achievement on measures of teacher-child interaction quality that target language and literacy interactions (i.e., ELLCO domains) showed strong relationships to children's criterion language and literacy achievement, and change in achievement. Teachers' change in performance on global teacher-child interaction quality measures (i.e., CLASS domains) significantly predicted three and four-year olds' norm-referenced language and literacy achievement. Previous studies examining such relationships did not use multiple measures to evaluate teacher quality, and looked at point-in-time ratings of quality on the measures they employed (Burchinal et al., 2009; Curby et al., 2009; Early et al., 2006; Early et al., 2007; Howes et al., 2008; Mashburn et al., 2008). Use of the CLASS and ELLCO in the current study demonstrated differing relationships to norm-referenced versus criterion measures, which may provide important information about key instructional practices that influence each type of student assessment. Analysis of change ratings may provide additional information on differences in students' classroom experiences over time, and help to inform professional development efforts and measure effectiveness of professional development.

Teacher quality assessment tools. The data for this study included the new CLASS Classroom Organization domain, which was not included for analysis in any of the research reviewed for this dissertation. The results of this study suggest that students who begin the year in classrooms with lower levels of classroom organization, as measured by the CLASS, demonstrate significantly lower achievement than their peers

who begin in classrooms rated higher on the CLASS, despite their teachers' positive change in rating by the end of the year.

Likewise, the revised ELLCO tool has not been used in similar research. Data from this study demonstrate that students who begin the year in classrooms with teachers who provided lower quality print and early writing environments, as evidenced by the ELLCO Print & Early Writing domain, make fewer gains from fall to spring on the criterion language measure than their peers who begin in classrooms with higher quality print and early writing environments. This may indicate that change in performance is less important than teachers' starting point on certain measures.

In summary, the results of this study suggest that peer characteristics are more predictive of students' achievement on criterion measures than norm-referenced measures. Furthermore, results from this study indicate that teachers' change in quality over time acts as a moderator for student achievement on norm-referenced and criterion measures, but different teacher quality assessments may be needed to analyze these relationships fully. Finally, teachers' change in quality in print and writing environment and classroom organization may be less important than initial quality.

Potential Implications

Results from this research have potential policy, practice and research implications.

Policy implications. This and similar research may support expansion of funding for targeted initiatives that address those variables most predictive of student achievement in language and literacy. This research lends some support to previous findings that ongoing professional development impacts teacher practice (Neuman &

Cunningham, 2009). Funding for professional development aimed at changing the quality of teacher-child interactions may support teachers' change in quality, and impact student achievement in language and literacy.

Practice implications. Findings from this research may support other similarly funded programs in evaluating program effectiveness for various student groups (e.g., children of different gender, race/ethnicity, SES). Findings from this research have direct implications for teacher professional development efforts, highlighting areas most related to student achievement on norm- and criterion-referenced language and literacy assessments used throughout the nation. Results from this study may provide some guidance school administrators by informing key skills and attributes of teacher quality to target when making hiring decisions and crafting teacher evaluation systems, Furthermore, outcomes of this research help to inform best practices for schools related to classroom peer and instruction characteristics most predictive of student achievement on norm- and criterion-referenced language and literacy measures. Additional analyses analyzing possible interaction effects between classroom-level variables are warranted.

Research implications. This researcher is aware of no published studies using HLM that analyze how the average change in quality of teacher-child interactions across a school year relates to students' language and literacy development. The studies reviewed for this dissertation looked at teacher ratings at one time in the school year. In addition, this research included peer characteristics not analyzed in larger studies using validated teacher quality tools, specifically number of ELL and special education students per class. Only one of the reviewed studies used the number of minority students per class as a control variable, but results were not reported to inform potential effects of peer

characteristics (Early et al., 2007). Future research should address the identified study limitations and expand on the scope of the research presented in this dissertation, such as focus on possible interaction effects between classroom-level variables.

Areas for future investigation include examining growth models on the classroom quality and student progress monitoring data. Because this research looks at change rather than growth, it is not clear at what point in the year teachers' ratings improved on this domain. Average ratings across the school year, or a growth model using all three time points of CLASS and ELLCO data, may have yielded additional more helpful information about how teacher-child interaction quality ratings. Child criterion measures might have yielded more consistent and conclusive information with the inclusion of growth analysis, as well, since children are administered the progress monitoring measures at three time points in the year.

In addition, potential significant interactions between independent variables were not examined due to power concerns based on the large number of variables already used in the study. It is possible that the teacher quality variables interact differently when examined as interactions with peer characteristics such number of ELL or special education students (Gallagher & Lambert, 2006) or variables created to approximate school-level differences such as the number of daily small groups (Early et al., 2007). There is also evidence with slightly older children that beginning of year quality and focus of teachers' instructional interactions differentially supports student achievement based on entering student achievement (Connor, Morrison & Katch, 2004; Connor, Son, Hindman & Morrison, 2005; Connor, Piasta, Fishma, Glasney, Schatschneider, Crowe et al., 2009). Future research looking at these and other potential interaction relationships

will help to expand the current research on relationships between student achievement and peer and classroom quality characteristics.

Furthermore, limited information on the history of services and out-of-school influences on students. Specifically, data were not available to analyze whether students' participation in special subjects (e.g., foreign language or music) or other learning opportunities provided by individual schools (e.g., after school programs, theater programs) explained any variance in children's language and literacy development. Participation in special subjects or other school activities may have affected students' language and literacy development. Students' home exposure to high quality language and literacy environments is known to be strongly and positively related to student achievement (Hart & Risley, 1985; National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN], 2004). Inclusion of data estimating home support for language and literacy practices might have influenced these results.

With the accountability inherent under the No Child Left Behind (NCLB) act, schools are incentivized to begin working with students as early as possible to support their academic achievement. However, more guidance needs to be provided to outline what quality, effective implementation of UPK should look like, and which components of high quality implementation are most effective for different at risk student groups. Several states have begun this work, and the research on effectiveness is quite promising. Analysis of teacher, classroom peer and student variables most predictive of positive academic outcomes in language and literacy, such as the analyses performed in the present study, can aid policy makers and practitioners, as well as providing guidance

about possible targeted areas of future experimental research.

Limitations

There are several limitations to the dissertation research. First, the study is limited by possible selection bias in that the study included only students within public charter schools; parents self-select students for enrollment, so final enrollment may not be representative of traditional public school enrollment in the city, or other charter schools in the city. Because of the non-experimental design, no inferences can be drawn about cause and effect. However, the use of pre-test data on norm-referenced assessments does allow for comparison to the normative sample from which the assessments were derived (Shadish et al., 2002). Also, the use of assessments that employed standardized administration protocols allows for secondary data comparison to other research using the same assessments with similar samples (Shadish et al., 2002). Analysis of single program year effect size differences for the assessment (i.e., PPVT) used in the current study and literature review studies indicate that on average, children in DCPEL made more gains in receptive language than children in any of the previous studies. Likewise, teacher quality as measured by the CLASS Emotional Support and Instructional Support was higher for DCPEL teachers, on average, than for teachers observed in the previous research.

This study did not evaluate classroom-level fidelity of implementation to curricula, or take into account professional development or teacher evaluation variables at individual schools. Fidelity of implementation of the core curricula, school-based professional development provided outside of the grant, and teacher evaluation variables may have differentially impacted teacher and student performance beyond the identified variables.

In addition, this study did not control for baseline teacher performance on the quality variables. Baseline quality likely impacted the amount of change demonstrated in each domain. Inclusion of baseline quality as a classroom-level control variable may have yielded different results. A final important area of consideration in this research is the limited range demonstrated on the classroom quality measures. In comparison to other studies, teachers in the DCPEL project have high quality classrooms, and there is limited variability on the majority of CLASS and ELLCO domains. Combined with the small number of classrooms, this may have masked potential effects of rating differences.

Conclusions

This study allowed classroom-level analyses of additional student and classroom peer variables that have been demonstrated to impact student achievement in language and literacy, including three-year old children, special education and ELL students, and classrooms housed in charter schools. The study also added to the limited body of existing research on enhancing student achievement in early childhood programs by analyzing universal preschool classrooms serving both three and four-year old children. This may be the first study analyzing these relationships with young children in the rapidly expanding charter school arena. In addition, there are important policy and practice implications related to funding, research, program evaluation, professional development and best practices.

This research addressed a gap in the literature related to the relationship between quality of teacher-child interactions and outcomes for special education and ELL students in general education classrooms. Furthermore, the majority of studies completed on the efficacy of the CLASS, and all of the studies using the ELLCO, used prior versions of the

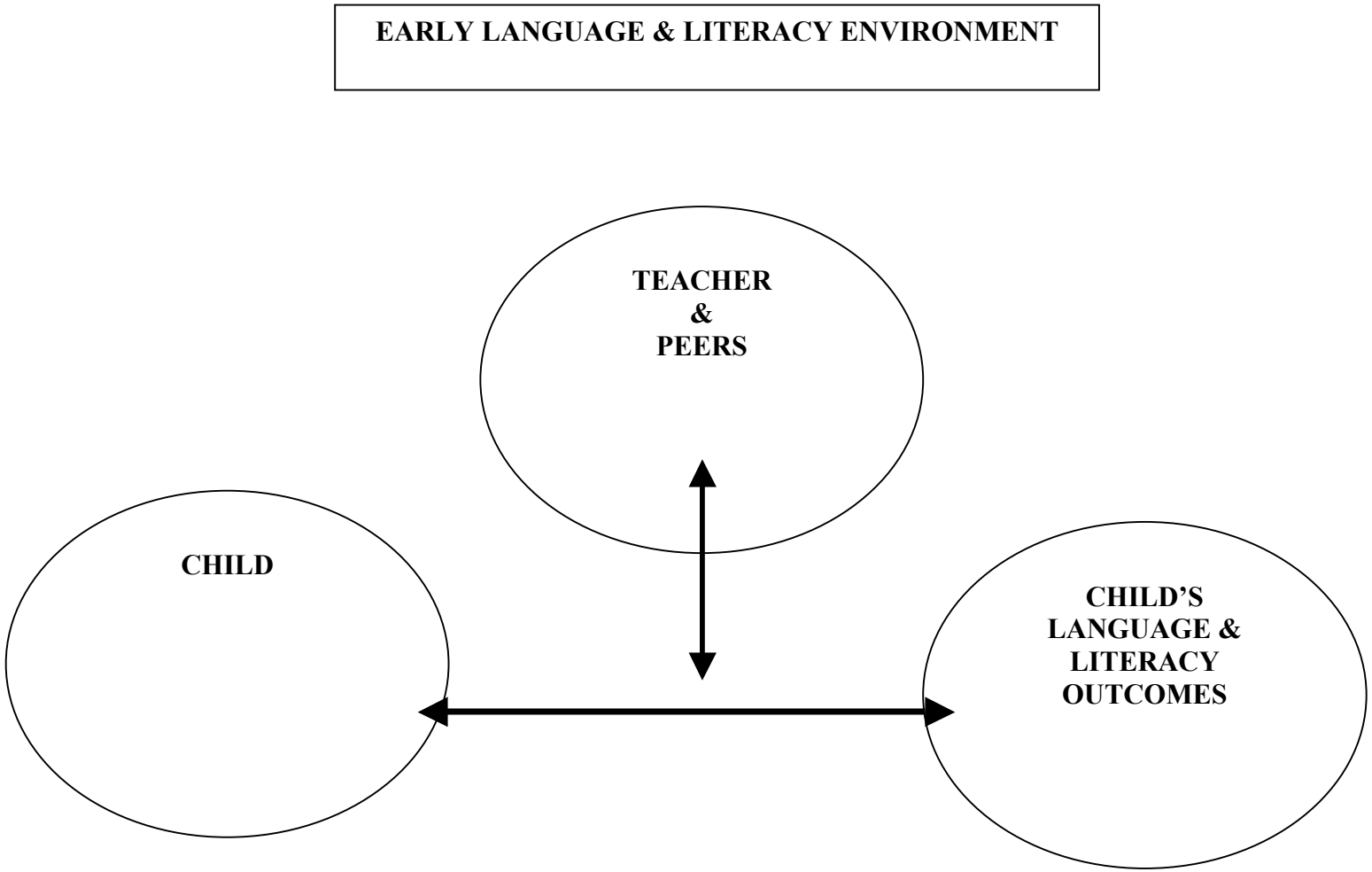
observational assessments. The study also provided preliminary evidence that classroom peer characteristics and change in teacher quality help predict norm- and criterion-referenced language and literacy achievement for three and four-year old children in public charter schools.

Table 1.1. *Seminal Research Comparison*

	High Scope/ Perry	Abecedarian	Chicago Parent-Child
Year Implemented	1962	1972	1983
Students Served	123 3- & 4-year old AA students	111 infants, 98% AA	1539, 95% AA, 5% Hispanic
Risk Factors	SES, Low IQ	SES	Low SES
Teacher-Student Ratio	1:6	1:3 birth to 2 1:6 3 to 5 yo	1:8
Lead Teacher Qualifications	BA in education field	Para+ infants MA+ preschool	BA+, ECE certification
Treatment Group	Academic-Social 2.5 hrs/day, 30 wks/year, Home Visits	Academic-Social 10 hrs/day, 50 wks/year, Home Visits	Academic-Social School Day, School Year, Parent Volunteer
Control Group	No program	Medical & Nutrition Only	Enrolled in public K
Longitudinal Data to:	40 years old	21 years old	23-24 years old
Last Follow-up	2005	2002	2007
Funding Source	Federal, Private	State, Federal, Private	Federal Title I
Return on Investment	$2/dollar spent	$4/dollar spent	$8.47/dollar spent
Gains	IQ, Vocabulary, Achievement, College, HS GPA, SPED, Adult literacy, Salary, Public Assistance, Home Ownership, Arrests	IQ, VIQ, Math/Reading Achievement, HS completion, College attendance, Parental age, Drug/cigarette use	HS completion, College, Conviction/ Incarceration, Public Assistance, Health Insurance, Depression

Note. From Promising Practices Network on Children, Families and Communities, 2009, RAND Corporation.

Figure 1. Adults and environment as mediators of children's understandings in early language and literacy activities



Adapted from *Engaging children with print: Building early literacy skills through quality read-alouds*. L.M. Justice & A. E. Sofka. NY: Guilford Press.

TABLE 2.1. *Content Summary of Studies*

Study Authors	Stated Purpose of Research	Participants	Design	Ind. Variables	Dep. Variables	Results	Limitations
Brown, Molfese, & Molfese, (2008)	Assess how teachers' educational attainment, experience and beliefs impact the development of letter identification	138 4-year old children from low income families attending public pre-K programs across 8 classrooms in 3 schools	Correlational (Multiple Regression)	Teacher education; Teacher experience; Teacher beliefs on importance of early literacy learning	Wide Range Achievement Tests (WRAT) Letter Knowledge Subscale	Teacher education and experience moderately correlated with child literacy outcomes Teacher beliefs not significantly correlated with child literacy outcomes	Small sample size of teachers; No observation of teacher behaviors, used only questionnaire; No ELL or SPED students in sample; Information not provided about geographic region of study, year study was conducted, length of school day, curricula used, teacher/child ratio
Burchinal, Howes, Pianta, Bryant, Early, Clifford & Barbarin (2008)	Assess the relationship between classroom quality and student outcomes at end of pre-kindergarten and kindergarten years	Secondary data analysis from NCEDL Multi-State program evaluation study	Correlational (Multiple Regression)	Teacher quality measured by Classroom Assessment Scoring System (CLASS) domains and Early Childhood Environmental Rating Scale – Revised (ECERS-R) items	Peabody Picture Vocabulary Test-III (PPVT-III); Oral & Written Language Scale (OWLS) Oral Expression Scale; Woodcok-Johnson III Letter Identification Subtest; Comprehensive Test of Phonologial Processing (CTOPP) Elision Subtest		
Curby, LoCasale-Crouch, Konold, Pianta, Howes, Burchinal et al. (2009)	Assess the relationship between pre-K quality profiles of teacher-child interactions and children's achievement gains	Secondary data analysis from NCEDL Multi-State/SWEEP program evaluations	Correlational (HLM)	Teacher quality assessed by CLASS dimensions	PPVT-III	Children in classrooms with the highest level of Instructional Support achieved higher language gains Children in classrooms with the highest level of Concept Development achieved the highest language gains	Sample had fewer children with low SES or low maternal education levels; No analysis of growth – fall scores controlled; Teacher quality assessed only in spring; No information regarding curricula used, length of school day for all students

TABLE 2.1 (Continued)

Content Summary of Studies

Study Authors	Stated Purpose of Research	Participants	Design	Ind. Variables	Dep. Variables	Results	Limitations
Early, Bryant, Pianta, Clifford, Burchinal, Ritchie et al. (2006)	Assess if teachers' education and or credentials predict classroom quality or children's academic gains	Secondary data analysis from NCEDL Multi-State/SWEEP program evaluations	Correlational (ANCOVA)	Teacher education; Teacher credentials;	Teacher quality measured by CLASS domains and ECERS-R items PPVT-III; OWLS; WJ-III Sound Awareness; Researcher-Created Letter ID test	Teacher education predictive of child math gains CDA credential predictive of children's gains in basic skills Education, training and credentials not consistently related to classroom quality or other academic gains	Classroom quality measured only in fall
Early, Maxell, Burchinal, Ebanks, Henry et al. (2007)	Assess the relationship between teachers' educational degree and major and classroom quality and student academic outcomes	Secondary data analysis from seven early childhood program evalution studies, including FACES and NCEDL Multi-State/SWEEP	Correlational (HLM)	Teacher education measured by highest level of education and major	Classroom quality measured by ECERS-R; PPVT-III; WJ-III Letter-Word Identification	6 of 27 analyses yielded positive significant correlations between teacher education and student outcomes; 2 of 27 analyses yielded negative significant correlations Weak correlation between teachers' highest degree and student language outcomes; no differences for degree type Significant correlation between teachers' highest degree and Letter ID at two sites	
Gallagher & Lambert (2006)	Assess if concentration of children with special needs in general education classrooms is associated with child outcomes and teacher quality	600 children in preschool and pre-K across 70 Head Start classrooms	Correlational (HLM)	Teacher quality measured by Assessment Profile for Early Childhood Programs: Research Edition II	Metropolitan Early Childhood Assessment Program Pre-literacy Inventory Print Concepts and Story Retelling subtests	No main effect on child outcomes related to concentration of children with special needs; Classrooms with larger concentrations of children with special needs were rated higher in classroom quality; Children in high quality classrooms with greater than 20% children with special needs scored lower in print concepts	Relied on parent report of special needs status; No pretest administered; Does not address curricula

TABLE 2.1 (Continued)

Content Summary of Studies

Study Authors	Stated Purpose of Research	Participants	Design	Ind. Variables	Dep. Variables	Results	Limitations
Gerde & Powell (2009)	Assess if teachers' formal education and professional training in early childhood related to use of more academic versus behavior-focused utterances during book readings and children's growth in receptive language	60 Head Start teachers and 341 3-5 year old children.	Correlational (Linear Growth Modeling)	Teacher education measured by self-report	Teacher verbal book-reading behavior measured by researcher created observational tool; PPVT-III	Level of teachers' early childhood education related to initial use of more book-focused language, growth in academic-focused language use, and outcome levels of academic language use. Children in classrooms where teachers used more book-focused language had higher gains in receptive vocabulary. Teachers with more ECE education produced fewer behavior-related utterances	Focused only on receptive vocabulary using one measure; No information about time between pre-/posttest; No information about classroom characteristics that might have impacted teacher utterances, such as initial PPVT means
Howes, Burchinal, Pianta, Bryant, Early, Clifford et al. (2008)	Assess the relationship between pre-K quality profiles of teacher-child interactions and children's achievement gains	Secondary data analysis from NCEDL Multi-State/SWEEP program evaluations	Correlational (HLM)	Teacher quality assessed by CLASS domains and ECERS-R items	PPVT-III; OWLS; Identifying Letters	Children showed larger gains in academic outcomes when they experienced higher-quality instruction or closer teacher–child relationships.; Gains were not related to characteristics of the child or program (i.e., ratio, teacher qualifications and program location and length).	6 months between student assessments; Varying levels of children's prior program experience, from 1 to 3 years; Does not address gains for ELL students; Did not use program characteristics in HLM model to assess classroom-level differences Does not address curricula
Mashburn, Pianta, Hamre, Downer, Barbarin, Bryant et al. (2008)	Assess association between child outcomes and pre-K program quality measured by NIEER, pre-K classroom quality measured by ECER-R, and quality of teacher-child interaction measured by CLASS IS	Secondary data analysis from NCEDL Multi-State/Sweep program evaluations	Correlational (HLM)	School program quality; Teacher quality assessed by CLASS domain and ECERS-R items	PPVT-III; OWLS; WJ-III Rhyming; Letter Naming	Classroom quality measured by ECERS-R associated with children's expressive language outcomes Teachers' Instructional Support score positively correlated with all measures of academic and language outcomes	Did not measure teacher qualifications or PD participation; Evaluated only pre-K students; No information on differences between children whose parents gave consent and those who did not; Does not address curricula

TABLE 2.1 (Continued)

Content Summary of Studies

Study Authors	Stated Purpose of Research	Participants	Design	Ind. Variables	Dep. Variables	Results	Limitations
Missall, McConnell, & Cadigan (2006)	Assess the variance in literacy skill growth for different groups of students Assess the variance in literacy skill growth for different groups of students based on classroom and teacher interaction characteristics	69 pre-K children divided into 4 groups: 26 with mild disabilities; 12 in Head Start with low SES; 19 ELLs; 12 in family education programs who were not low SES, ELL or identified as having a disability	Correlational (HLM)	Percent time spent in instructional situations measured by Ecobehavioral System for the Complex Assessment of Preschool Environments (ESCAPE)	Individual Growth and Development Indicators (IGDI) Picture Naming, Rhyming, and Alliteration subtests	All children showed growth across measures Correlations between subtest and ESCAPE category differed for the four groups ECFE students scored highest and had largest slopes across measures, though rate of growth was similar for ELL and SLD children in Picture Naming and Rhyming	Small sample and subsamples; Different groups sometimes had same instruction; All children not administered same number of assessments; Did not measure teacher qualifications or PD participation; Does not address program characteristics such as length of school day and curricula implemented Does not address curricula

Table 2.2. *Ecological Factors Considered by Study*

		Brown et al. (2008)	Burchinal et al. (2008)	Curby et al. (2009)	Early et al. (2006)	Early et al. (2007)	Gallagher & Lambert (2006)	Gerde & Powell (2006)	Howes et al. (2008)	Mashburn et al. (2008)	Missall et al. (2006)
Student Factors	Race/ethnicity		x	x		x			x	x	
	Gender		x	x		x	x		x	x	
	SES background	x	x	x		x			x	x	x
	Maternal education		x	x		x	x		x	x	
	ELL			x							x
	SPED						x				x
	Age						x	x	x		
	Fall performance	x	x	x			x	x	x	x	x
Teacher Factors	Education level	x			x	x		x	x	x	
	Field of study				x	x		x		x	
	Years teaching	x									
	Credential				x						
	Quality		x	x	x	x	x	x	x	x	x
	Race/ethnicity					x					
Classroom Factors	Class size					x	x			x	
	Teacher to student ratio					x	x		x	x	
	Comprehensive curriculum									x	
	School day/ hours per week		x		x				x		
	Setting				x	x			x		
	Mean student factors				x	x	x		x		

Table 2.3. *Descriptive Data by Study*

		Brown et al. (2008)	Burchinal et al. (2008)	Curby et al. (2009)	Early et al. (2006)	Gallagher & Lambert (2006)	Gerde & Powell (2006)	Howes et al. (2008)	Mashburn et al. (2008)	Missall et al. (2006)
Student Factors	Total Students	138	759	2028	939	600	341	844	2439	69
	% White	*NI*	43	49	41	28	40	42	46	25
	% Black	*NI*	23	21	24	67	39	24	21	23
	% Hispanic	*NI*	24	14	25	4	17	24	17	29
	% Other Ethnicity	*NI*	10	16	10	1	4	10	15	23
	% Boys	54	48	48	49	49	52	49	49	59
	% ELL	0	25	*NI*	11	*NI*	17	*NI*	12	28
	% SPED	0	0	0	0	20	*NI*	0	0	38
	Mean Age	49.86	54.24	56.00	54.7	58.95	55.00	56.00	*NI*	58.36
	% Low SES	100	56	58	57	100	100	58	47	17
	Mean Maternal Education	12.42	12.60	12.8	*NI*	12.55	*NI*	12.63	12.9	*NI*
Teacher Factors	% BA/BS	50	*NI*	71.3	59	*NI*	42	71.3	70	*NI*
	% ECE Focus	50	*NI*	60.2	64	*NI*	63	60.2	60	*NI*
	Mean Years teaching	9.63	*NI*	9.00	9.69	*NI*	9.12	9.00	*NI*	*NI*
	% ECE Credential	*NI*	*NI*	*NI*	75	*NI*	*NI*	*NI*	*NI*	*NI*
	% White	*NI*	*NI*	71	61	*NI*	*NI*	71	*NI*	*NI*
	% Black	*NI*	*NI*	11	19	*NI*	*NI*	11	*NI*	*NI*
	% Hispanic	*NI*	*NI*	16	14	*NI*	*NI*	16	*NI*	*NI*
Classroom Factors	Mean Class size	*NI*	*NI*	19	*NI*	16.60	*NI*	19	*NI*	*NI*
	Teacher to student ratio	*NI*	*NI*	1:8.6	1:6.88	1:7.40	1:9.10	1:8.6	1:10	*NI*
	% Comprehensive curriculum	*NI*	*NI*	*NI*	*NI*	*NI*	*NI*	*NI*	57	*NI*
	% Full Day	*NI*	*NI*	*NI*	51	*NI*	*NI*	48	*NI*	*NI*
	Mean Hours/day	*NI*	*NI*	4.9	*NI*	*NI*	*NI*	5.09	*NI*	*NI*
	% in school	*NI*	*NI*	62	54	*NI*	*NI*	55	*NI*	*NI*
	% Head Start	*NI*	*NI*	15.1	*NI*	100	100	*NI*	*NI*	*NI*

Note: NI – No information provided. Early et al. (2007) not included due to variation in data across analyzed program evaluation studies.

Table 2.4 *Mean effect sizes for dependent variables across literature review studies*

	WRAT T1	(SD)	WRAT T2	(SD)	***Cohen's d***
Brown et al. (2008)	2.96	4.52	6.84	5.74	**0.76**

	PPVT T1	(SD)	PPVT T2	(SD)	***Cohen's d***	OWLS T1	T1 SD	OWLS T2	(SD)	***Cohen's d***	WJ T1	T1 SD	WJ T2	(SD)	***Cohen's d***	WJ-R T1	T1 SD	WJ-R T2	(SD)	***Cohen's d***
Burchinal et al. (2008)	93.15	13.71	97.71	12.49	**0.35**	91.46	12.16	95.16	12.91	**0.30**	1.66	2.62	2.25	4.53	**0.17**	2.13	0.82	7.25	4.53	**1.91**

	PPVT T1	(SD)	PPVT T2	(SD)	***Cohen's d***
Curby et al. (2009)	94.58	14.86	97.43	13.97	**0.20**

	PPVT T1	(SD)	PPVT T2	(SD)	***Cohen's d***	OWLS T1	T1 SD	OWLS T2	(SD)	***Cohen's d***	WJ T1	T1 SD	WJ T2	(SD)	***Cohen's d***	Letters T1	T1 SD	Letters T2	(SD)	***Cohen's d***
Early et al. (2006)	92.64	13.89	95.69	13.58	**0.22**	91.28	12.20	94.79	12.29	**0.29**	1.63	2.57	2.95	3.54	**0.43**	7.26	8.57	12.26	9.50	**0.55**

Gallagher & Lambert (2006)	Not able to report

Table 2.4 cont.

	PPVT T1	(SD)	PPVT T2	(SD)	***Cohen's d***
Gerde & Powell (2009)	88.15	15.72	91.71	14.45	**0.24**

	PPVT T1	(SD)	PPVT T2	(SD)	***Cohen's d***	OWLS T1	(SD)	OWLS PC T2	(SD)	***Cohen's d***	Letters T1	(SD)	Letters T2	(SD)	***Cohen's d***
Howes et al. (2008)	93.06	13.59	95.96	13.57	**0.21**	91.51	12.14	95.06	12.26	**0.29**	7.31	8.62	12.31	9.54	**0.55**

	PPVT T1	(SD)	PPVT T2	(SD)	***Cohen's d***	OWLS T1	(SD)	OWLS PC T2	(SD)	***Cohen's d***	WJ-R T1	(SD)	WJ-R T2	(SD)	***Cohen's d***	Letters T1	(SD)	Letters T2	(SD)	***Cohen's d***
Mashburn et al. (2008)	94.20	15.00	96.30	14.30	**0.14**	91.60	13.10	93.60	13.00	**0.15**	2.26	3.23	3.65	4.02	**0.38**	8.69	9.03	13.90	9.42	**0.56**

Missall et al. (2006) Not able to report

	FACES		**GECS**		**MAF**		**NCEDL**		**PCER**	
Early et al. (2007)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
PPVT Time 1	84.48	(11.01)	92.38	(15.90)	85.42	(17.91)	94.00	(15.01)	89.83	(15.61)
PPVT Time 2	86.19	(11.68)	96.54	(14.50)	89.57	(16.20)	96.29	(14.31)	93.78	(14.42)
Cohen's d		0.15		0.27		0.24		0.16		0.26

Table 2.5. *Definitions for Internal, Construct, Statistical Conclusion, and External Validity Criteria*

Criterion	Definition
Internal Validity Criteria	
Selection Bias	Between group differences with selected participants that may contribute to differential outcomes.
Ambiguous Temporal Precedence	Uncertainty about order of introduction of dependent and independent variables that make it difficult to ascertain which variable was cause or effect.
Testing	Repeated exposure to an assessment measure may impact subsequent scores, making it difficult to ascertain effect of independent variable.
Construct Validity	
Mono Operation Bias	Using only one dependent variable to measure a construct might not adequately measure the intended construct or measure unintended constructs.
Confounding of Constructs	Failure to describe all constructs may result in invalid inferences.
Statistical Conclusion Validity	
Restriction of Range	Reduced range on a variable typically weakens the relationship between it and another variable.
Unreliability of Treatment Implementation	Measurement error weakens the relationship between two variables and strengthens or weakens the relationships among three or more variables.
External Validity	
Restricted Generalizability of Participants	A relationship determined with one sample might not transfer to a different sample
Restricted Generalizability of Treatment	The intervention may be too specialized or inadequately explained to generalize or apply to "real world" settings.
Context-Dependent Mediators	A correlational mediator in one context or setting might not predict outcomes in another context

Note. Adapted from Shadish, Cook & Campbell, 2002.

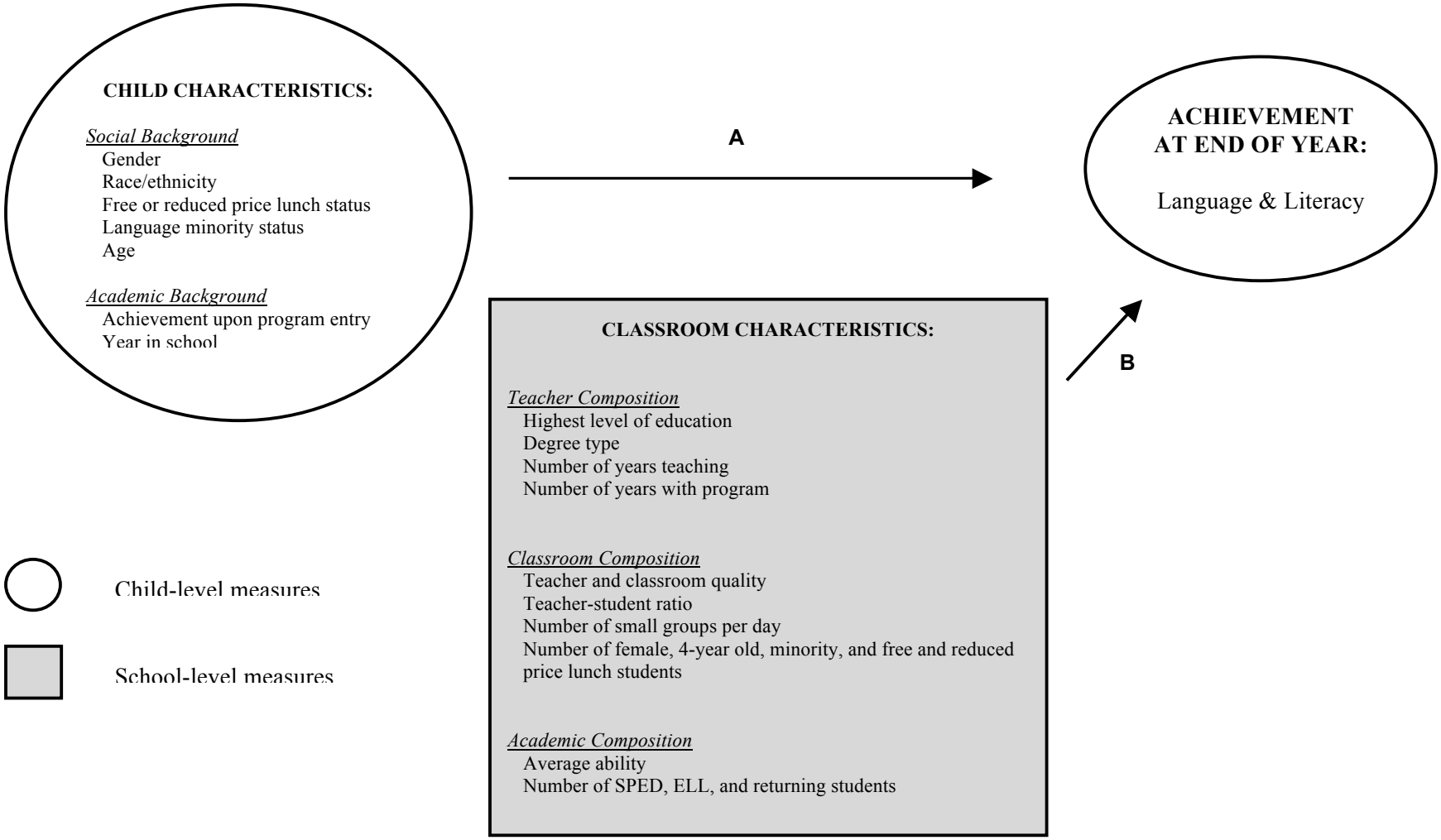


FIG. 2. – Multilevel model for evaluating the effects of child and classroom characteristics on children's learning in language and literacy

References

Abbott-Shim, M., & Sibley, A. (1998). *Assessment Profile for Early Childhood Programs-Research Edition II*. Atlanta, GA: Quality Assist.

Administration for Children and Families. (2006). "FACES Findings: New Research on Head Start Outcomes and Program Quality". Retrieved March 24, 2009, from http://www.acf.hhs.gov/programs/opre/hs/faces/reports/faces_findings_06/faces06_title.html.

Administration for Children and Families. (2009). "Head Start Program Fact Sheets". Viewed March 24, 2009, from http://www.acf.hhs.gov/programs/ohs/about/index.html#factsheet.

AppleTree Institute for Education Innovation (2007). Early Reading First project application: DC Partnership for Early Literacy. Available from http://www2.ed.gov/policy/gen/leg/foia/grants/erfgappletree.pdf

Barnett, W.S., Lamy, C., and Jung, K. (2005). *The effects of state pre-kindergarten programs on young children's school readiness in five states*. NIEER Policy Report. Retrieved March 18, 2009 from http://nieer.org/docs/index.php?DocID=129.

Bloom, H.S., Richburg-Hayes, L., & Black, A. R. (2005). Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions. New York: MDRC. Available at: http://www.mdrc.org/publications/417/full.pdf

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*(7), 513-531.

Bronfenbrenner, U., & Morris, P.A. (1998). The ecology of development processes. In W. Damon, & R.M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th edition, pp. 993-1029). New York: John Wiley & Sons.

Brown, E., Molfese, V., & Molfese, P. (2008). Preschool student learning in literacy and mathematics: Impact of teacher experience, qualifications, and beliefs on an at-risk sample. Journal of Education for Students Placed at Risk, 13(1), 106-126.

Bryk, A.S. and Raudenbush, S.W. (1992). *Hierarchical Linear Models.* Newbury Park, CA: Sage Publications, Inc.

Burchinal, M., Howes, C., Pianta, R. C., Bryant, D., Early, D., Clifford, R. et al. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Developmental Science, 12,* 140–153.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology, 37*, 231-242.

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J.J, & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6,* 42-57.

Capizzi, A. M., & Fuchs, L. S. (2005). Effects of curriculum-based measurement with and without diagnostic feedback on teacher planning. Remedial & Special Education, 26, 159–174.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scale (OWL).* Circle Pines, MN: American Guidance Service.

Christina, R., and Nicholson-Goodman, J. (2005). Going to Scale with High-Quality Early Education: Choices and consequences in universal pre-kindergarten efforts. Santa Monica, CA: RAND. [Online]. Available: http://www.rand.org/publications/TR/TR237

Connor, C. M., Morrison, F. J., & Katch, E. L. (2004a). Beyond the reading wars: The effect of classroom instruction by child interactions on early reading. *Scientific Studies of Reading, 8*, 305 – 336.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E. et al. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development, 80*, 77-100.

Connor, C. M., Son, S., Hindman, A., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology, 43*, 343 – 375.

Curby, T., LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M. et al. (2009). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. *Early Education and Development, 20(*2), 346-372.

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15(3),* 358-374.

Deno, S.L., Marston, D.B., & Tindal, G. (1986). Direct and frequent curriculum based measurement: An alternative for educational decision making. *Special Services in the Schools, 2*(2-3), 5-28.

District of Columbia Public Schools [DCPS]. (2011). Facts and figures. Retrieved February 6, 2011 from k12.dcps.org.

Dotterer, A. M., Burchinal, M., Bryant, D., Early, D., & Pianta, R. (2009). Comparing universal and targeted pre-K programs. In R. Pianta & C. Howes (Eds.), *The promise of pre-kindergarten*. Brookes Publishing Co., Inc.

Dunn, L.M., & Dunn, L.M. (1997). *Peabody Picture Vocabulary Test* (3rd Ed.). Circle Pines, MN: American Guidance Service.

Early, D. M., Bryant, D., Pianta, R., Clifford, R., Burchinal, M., Ritchie, S. et al. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre- kindergarten? *Early Childhood Research Quarterly, 21*, 174 – 195.

Early, D., Barbarin, O, Bryant, D., Burchinal, M., Chang, F., Clifford, R., Crawford, G., Weaver, W., Howes, C., Ritchie, S., Kraft-Sayre, M., Pianta, R., & Barnett, W. S. (2005). Pre-Kindergarten in Eleven States: NCEDL's Multi-State Study of Pre-Kindergarten & Study of State-Wide Early Education Programs (SWEEP). NCEDL Working Paper. Chapel Hill: The University of North Carolina, FPG Child Development Institute, NCEDL. Retrieved May 1, 2009 from http://www.fpg.unc.edu/~ncedl/pdfs/SWEEP_MS_summary_final.pdf

Early, D., Maxwell, K., Burchinal, M., Bender, R., Ebanks, C., Henry, G. et al. (2007). Teachers' education, classroom quality, and young children's academic skills:

Results from seven studies of preschool programs. *Child Development, 78*(2), 558-580.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Finn, J.D. & Achilles, C.M. (1999). Tennessee's class size study: Findings, implications, misconceptions. Educational Evaluation and Policy Analysis, 21(2), 97-109.

Frank Porter Graham Child Development Institute. (2005, Spring). *Early Developments*. Chapel Hill: The University of North Carolina, FPG Child Development Institute, NCEDL.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991). Effects of expert system advice within curriculum-based measure- ment on teacher planning and student achievement in spelling. School Psychology Review, 20, 49–66.

Gallagher, P., & Lambert, R. (2006). Classroom Quality, Concentration of Children With Special Needs, and Child Outcomes in Head Start. *Exceptional Children, 73*(1), 31-52.

Gerde, H., K., and Powell, D. R. (2009). Teacher education, book-reading practices, and children's language growth across one year of Head Start. *Early Education & Development, 20(*2), 211 — 237.

Greenwood, C. R., Carta, J. J., Kamps, D., & Delquadri, J. (1997). Ecobehavioral assessment systems software. Lawrence, KS: Juniper Gardens Children's Project, University of Kansas.

Hanushek, E.A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143-163.

Harms, T., Clifford, R.M., & Cryer, D. (1998). *Early childhood environment rating scale* (Rev. ed). New York: Teachers College Press.

Hart, B., & Risley, R.T. (1995). *Meaningful differences in the everyday experience of*

Harwell M. R., Maeda Y., Lee K. (2004, April). *Replicating and extending White's (1982) Meta-analysis of the relationship between SES and student achievement*. *Paper presented at the annual meeting of the American Educational Research Association*, San Diego, CA.

Head Start Family and Child Experiences Survey (FACES) (1998). *Color bears*. Washington, DC: Westat.

Henry, G. T., Henderson, L. W., Ponder, B. D., Gordon, C. S., Mashburn, A. J., & Rickman, D. K. (2003). *Report of the findings from the Early Childhood Study: 2001– 02*. Atlanta: Georgia State University, Andrew Young School of Policy Studies.

Henry, G., Gordon, C., Mashburn, A., & Ponder, B. (2001). *Pre-K longitudinal study: Findings from the 1999–2000 school year*. Atlanta: Georgia State University, Applied Research Center.

Hoffmann, J. P. (2004). Generalized linear models: An applied approach. Boston, MA: Pearson Education.

Horowitz, S. (2006, November). Research Roundup, *National Center for Learning Disabilities,* Retrieved August 4, 2008 from

http://www.ncld.org/content/view/1109/480/.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R. et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*, 27-50.

Joyce, B. & Showers, J. (1982). The coaching of teaching. *Educational Leadership*, *40*, 4-10.

Justice, L. M. and Sofka, A. E. (2010). *Engaging children with print: Building early literacy skills through quality read-alouds*. NY: Guilford Press.

Karoly, L., Ghosh-Dastidar, B., Zellman, G., Perlman, M., & Fernyhough, L. (2008). *Nature and quality of early care and education for California's preschool-age children: Results from the California Preschool Study*. RAND Corporation.

Kutner, N.N. (2004). *Applied Linear Regression Models*, 4th edition, McGraw-Hill: Irwin.

LaParo, K., Pianta, R., Hamre, B., & Stuhlman, M. (2002). CLASS: PK version. Unpublished measure, University of Virginia.

La Paro, K.M., Pianta, R.C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the pre-kindergarten year. *The Elementary School Journal, 104*(5), 409-426.

Lee, V. E., Burkam, D. T., Ready, D. D., Honigman, J., & Meisels, S. J. (2006). Full-day versus half-day kindergarten: In which program do children learn more? *American Journal of Education, 112*(2), 163–208.

Lennon, J. E., & Slesinski, C. (1999). Early intervention in reading: Results of a screening and intervention program for Kindergarten students. *School Psychology Review*, 28, 353-364.

Liu, X. , Spybrook, J., Congdon, R., Martinez, A., & Raudenbush, S.W., (2009). Optimal Design for Longitudinal and Multilevel Research V.2.0.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent- variable longitudinal study. *Developmental Psychology, 36,* 596–613.

Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D. et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. Child Development, 79(3), 732-749.

Massey, L., and Barnett, S. (2003). *A benefit cost analysis of the Abecedarian Early Childhood Intervention*. New Brunswick, NJ: National Institute for Early Education Research.

McBride-Chang, C. (1998). The development of invented spelling. *Early Education and Development, 9*, 147– 160.

McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson IV: Tests of Achievement*. Itasca, IL: Riverside Publishing.

Missall, K., McConnell, S., & Cadigan, K. (2006). Early literacy development: Skill growth and relations between classroom variables for preschool children. *Journal of Early Intervention, 29*(1), 1-21.

National Center for Early Development and Learning (2001). *Identifying Letters and Identifying Numbers.* Unpublished.

National Center for Educational Statistics (NCES, 2002). Early Childhood Longitudinal Study, *Kindergarten Study*, 2000.

National Institute of Child Health and Human Development Early Child Care Research Network. (2004). Multiple pathways to early academic achievement. *Harvard Educational Review, 74*, 1 – 29.

National Research Council. (1999). Starting out right: A guide for promoting children's reading success. Washington, DC: National Academy Press.

Neuman, S.B. (2006). The knowledge gap: Implications for early education. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research, Volume 2* (pp. 29-40). New York: Guilford Press.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy practices. *American Educational Research Journal, 46*(2), 532-566.

Nurss, J. A. (1995). *Metropolitan early childhood assessment program: Preliteracy inventory.* San Antonio, TX: The Psychological Corporation.(pp. 277-291). Baltimore, MD: Paul H, Brookes.

Peisner-Feinberg, E. S., & Maris, C. L. (2005a). Evaluation of the North Carolina More at Four Pre-kindergarten Program: Year 2 Report (July 1, 2002 – June 30, 2003). Chapel Hill, NC: FPG Child Development Institute. Retrieved October 20, 2009, from http://www.fpg.unc.edu/mafeval/pdfs/MAF_Yr2_Rpt.pdf

Peisner-Feinberg, E. S., & Maris, C. L. (2005b). Evaluation of the North Carolina More at Four Pre-kindergarten Program: Year 3 Report (July 1, 2003 – June 30, 2004). Chapel Hill, NC: FPG Child Development Institute. Retrieved October 20, 2009, from http://www.fpg.unc.edu/mafeval/pdfs/MAF_Yr3_Rpt.pdf

Perez-Johnson, I., and Maynard, R. (2007). The case for early, targeted interventions to prevent academic failure. *Peabody Journal of Education, 82*(4), 587-616.

Pianta, R.C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D. et al. (2005). Features of Pre-Kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*(3), 144-158.

Powell, D. R., Diamon, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology, 102*(2), 299-312.

Priest, J. S., McConnell, S. R., McEvoy, M. A., & Shin, J. (2000a). Early Childhood Research Institute on Measuring Growth and Development: Progress in five domains. Paper presented at the annual conference of the Division for Early Childhood, Council for Exceptional Children, Albuquerque, NM.

Priest, J. S., Silberglitt, B., Hall, S., & Estrem, T. L. (2000b). Progress on preschool IGDIs for early literacy. Presentation at Heartland Area Education Association, Des Moines, IA.

Ramey, C.T. and F.A. Campbell. (1984). Preventive education for high risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency, 88*(5), pp. 515-523.

Ramey, C. T., Ramey, S. L., Crowell, N. A., and Cathy, P. (2009). Evaluation of DC Partnership for Early Literacy: A project of AppleTree Institute for Innovation. *Early Reading First Grant # S359B080043.* Washington, DC: Georgetown University Center on Health and Education.

Raudenbush, S.W., & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analyses methods (2nd ed.), Thousand Oaks, CA: Sage.

RAND Corporation (n.d.). Promising Practices Network. Retrieved February 2010 from http://www.promisingpractices.net.

Reynolds, A. J., Ou, S. & Topitzes, J. W. (2004) Paths of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers. *Child Development, 75* (5), 1299-1338.

Reynolds, A. J., Temple, J. A., Robertson, D. L., and Mann, E. A. (2001a). *Age 21 Cost-Benefit Analysis of the Title I Chicago Child-Parent Center Program Executive Summary*. Retrieved March 24, 2009 from http://www.waisman.wisc.edu/cls/cbaexecsum4.html.

Reynolds, A., Temple, J., Robertson, D., & Mann, E. (2001b). Long-term

Sandoval-Martinez, 5. (1982). Findings from the Head Start bilingual curriculum development and evaluation effort. *NABE Journal*, 7, 1 - 12.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), Specific reading disability: A view of the spectrum (pp. 75–119). Timonium, MD: York Press.

Schickedanz, J. A., & Dickinson, D. K. (2005). *Opening the World of Learning: A comprehensive early literacy program*. Parsippany, NJ: Pearson Early Learning.

Shadish, W., Cook, T. & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin. Retrieved November 8, 2009, from http://courses.washington.edu/socw580/readings/.

Sirin S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75, 417–453.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, D.C.: National Academy Press.

Spybrook, J., Raudenbush, S.W., Congdon, R., & Martinez, A. (2009). Optimal Design for Longitudinal and Multilevel Research: Documentation for the Optimal Design Software V.2.0. Available at www.wtgrantfoundation.org.

StataCorp. 2009. *Stata statistical software: Release 11*. College Station, TX: StataCorp LP.

Stein M. L., Berends M., Fuchs D., McMaster K., Saenz L., Yen L., et al . (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis*, 30, 368–388.

Test, D.W., Fowler, C.H., Brewer, D.M., and Wood, W.M. (2005). A content and methodological review of self-advocacy intervention studies. *Exceptional Children, 72*, 101-125.

Torgesen, J.K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*, 7-26.

U.S. Department of Education, National Center for Education Statistics (NCES). (2002). Condition of Education. Washington, D.C.: NCES.

University of Minnesota ResearchWorks. (2005). Curriculum-based measurement (CBM): Student assessment. Retrieved June 2008 from http://www.cehd.umn.edu/research/highlights/CBM/.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development of validation of a process for screening referrals to special education. *School Psychology Review, 32*, 204–227.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-toremediate and readily remediated poor readers: More evidence against the IQachievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33*, 223–238.

Velluntino, F., Scanlon, D., & Tanzman, M. (1998). The case for early intervention in diagnosing reading disability. *Journal of School Psychology, 36,* 367-397.

Weikart, D. P., Deloria, D., Lawser, S., and Wiegerink, R. (1970). Longitudinal results of the Ypsilanti Perry Preschool Project (Monographs of the High/Scope Educational Research Foundation, 1). Ypsilanti, MI: High/Scope Press.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity," *Econometrica*, 48, 817-838.

Whitehurst, G.J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69, 848-872.

Whitehurst, G.J., & Lonigan, C. J. (2001). Emergent Literacy: Development from pre-readers to readers. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 11-29). New York: Guilford.

Whitehurst, G. J. (2001). *The NCLD Get Ready to Read Screen technical report*. New York: National Center for Learning Disabilities.

Wilkinson, G. (1994). *Wide Range Achievement Test-III.* Wilmington, DE: Jastak Associates. *young American children*. Baltimore: Brookes.

Yoon, K.S., Duncan, T., Lee, S.W., Scarloss, B., & Shapley, K.L. (2007). *Reviewing the Evidence on how Teacher Professional Development Affects Student Achievement.* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Zill, N., & Resnick, G. (2006). Emergent literacy of low-income children in head start: Relationships with child and family characteristics, program factors, and classroom quality. In S.B. Neuman & D.K. Dickinson (Eds.). *Handbook of Early Literacy Research* (pp. 347-371). New York: The Guilford Press.