

Effects of Preschool Curriculum Programs on School Readiness

Report from the Preschool Curriculum Evaluation Research Initiative



NCER 2008-2009
U.S. DEPARTMENT OF EDUCATION

 **NATIONAL CENTER FOR
EDUCATION RESEARCH**
Institute of Education Sciences

Effects of Preschool Curriculum Programs on School Readiness

Report from the Preschool Curriculum Evaluation Research Initiative

July 2008

Preschool Curriculum Evaluation Research Consortium

This report was prepared for the National Center for Education Research, Institute of Education Sciences, under contract numbers ED-01-CO-0039/0005 (Mathematica Policy Research, Inc.), ED-01-CO-0052/0004 (RTI International), and ED-04-CO0076/0007 (Synergy Enterprises, Inc.).

Disclaimer

The opinions and positions expressed in this report are the authors' and do not necessarily represent the opinions and positions of the Institute of Education Sciences or the U.S. Department of Education. Any references within the document to specific education products are illustrative and do not imply endorsement of these products to the exclusion of other products that are not referenced.

U.S. Department of Education

Margaret Spellings

Secretary

Institute of Education Sciences

Grover J. Whitehurst

Director

National Center for Education Research

Lynn Okagaki

Commissioner

July 2008

This report is in the public domain. While permission to reprint this publication is not necessary, the citation should be:

Preschool Curriculum Evaluation Research Consortium (2008). *Effects of Preschool Curriculum Programs on School Readiness* (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

This report is available for download on the IES website at <http://ncer.ed.gov>.

Alternative Formats

On request, this publication can be made available in alternative formats, such as Braille, large print, audiotape, or computer diskette. For more information, call the Alternative Format Center at (202) 205-8113.

Preschool Curriculum Evaluation Research Consortium¹

Institute of Education Sciences (IES)

Caroline Ebanks
Allen Ruby

Mathematica Policy Research (MPR), Inc.

John Love
Sarah Avellar
Steve Glazerman
Susan Sprachman

RTI International

Ina Wallace
Randall Bender
Renate Houts
Jun Liu
Melissa Raspa
Holly Rhodes

Florida State University

Christopher J. Lonigan
Christopher Schatschneider

Purdue University

Douglas Powell
Nancy File (University of Wisconsin-Milwaukee)

Success for All Foundation

Bette Chambers
Robert Slavin

University of California, Berkeley²

Prentice Starkey
Alice Klein
Douglas Clements (University at Buffalo,
State University of New York)
Julie Sarama (University at Buffalo,
State University of New York)

University of California, Berkeley³

Anne Cunningham
Marcia Davidson (RMC Research; University of Maine)

University of Missouri

Kathy R. Thornburg
Wayne Mayfield
Johnetta Morrison
Jacqueline Scott

University of North Carolina at Charlotte

Richard G. Lambert
Martha Abbott-Shim (Quality Counts)

University of New Hampshire

Jeffrey S. Priest
Leigh Zoellick

University of North Florida

Cheryl Fountain
Madelaine Cosgrove
Janice Wood

University of Texas Health Science Center at Houston

Susan Landry
Michael Assel
Susan Gunnewig
Paul Swank

University of Virginia

Laura Justice
Khara Pence
Alice Wiggins

Vanderbilt University

Dale Farran
Mark Lipsey

¹ The members of the Preschool Curriculum Evaluation Research (PCER) Consortium include the principal investigators and co-principal investigators from each of the 12 funded research projects, Institute of Education Sciences (IES) staff, and staff from RTI International (RTI) and Mathematica Policy Research (MPR), Inc., the evaluation contractors.

² This University of California, Berkeley research team partnered with researchers at the University at Buffalo, State University of New York to evaluate the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* curriculum in California and New York.

³ This University of California, Berkeley research team evaluated the *Ready, Set, Leap!* curriculum in New Jersey.

Page intentionally left blank.

Acknowledgments

The findings reported here are based on research conducted by the Preschool Curriculum Evaluation Research (PCER) program research teams, the evaluation contractors, and Institute of Education Sciences (IES) staff. This report is a product of the collaborative efforts of the PCER Consortium. The PCER Consortium consists of research teams from each participating grantee site, IES staff, and the evaluation contractors: RTI International (RTI) and Mathematica Policy Research (MPR), Inc. Appendix B of the report was authored by Randall Bender (RTI), Jun Liu (RTI), Ina Wallace (RTI), Melissa Raspa (RTI), and Margaret Burchinal (University of North Carolina at Chapel Hill).

The PCER Consortium would like to acknowledge Dr. Susan J. Kontos who served as the principal investigator for the *Project Approach* (Wisconsin) evaluation study from July 2002 to September 2003. Dr. Kontos, one of the country's leading researchers in early childhood education and care, died September 12, 2003.

We also acknowledge and thank Heidi Schweingruber and James Griffin for their contributions to the PCER program during their tenure at IES.

We are grateful to the schools, teachers, parents, and the children who participated in our assessments, interviews, and observations. Without their cooperation, none of this research would have been possible.

The listed authors represent only a part of the research team involved in this project. We would like to thank the research staff at each grantee site, especially the grantees' site coordinators who worked closely with the local preschool programs, kindergarten school staff, and the contractors' data collection teams to facilitate the successful collection of child-, parent-, teacher-, and classroom-level data. We appreciate the efforts of the classroom observers and the child assessors who were critical to the successful completion of data collection at each research site. We are also grateful to the many contractor staff members who have worked on data collection and data analysis tasks over the duration of the study.

The mention of trade names, commercial products, or organizations in the description of the projects, or the reporting of study findings, does not imply endorsement by the U.S. Government.

Page intentionally left blank.

Disclosure of Potential Conflicts of Interest

The PCER Consortium consists of research teams (principal investigators and co-principal investigators from each grantee site), IES staff, and the evaluation contractors, Mathematica Policy Research (MPR), Inc. and RTI International (RTI). Most of the grantee research teams, IES staff, and contractor staff from MPR and RTI have no interests that could be affected by findings from the evaluation of the curricula that are highlighted in this report.

It is important to note that four of the PCER initiative research teams developed curricula that were implemented at their respective research sites. The Success for All Foundation (SFA) developed the *Curiosity Corner* curriculum, which was implemented in preschool classrooms in Florida, Kansas, and New Jersey. Dr. Christopher Lonigan and his colleagues at Florida State University developed the *Literacy Express* curriculum, which was implemented in public pre-kindergarten classrooms in Florida. Drs. Prentice Starkey and Alice Klein are the developers of the *Pre-K Mathematics* curriculum. Drs. Douglas Clements and Julie Sarama are the developers of the *DLM Early Childhood Express Math software*. The *Pre-K Mathematics* curriculum and the *DLM Early Childhood Express Math software* were implemented jointly in Head Start and public pre-kindergarten classrooms in California and New York. Drs. Cheryl Fountain, Madelaine Cosgrove, and Janice Wood are on staff at the Florida Institute of Education, University of North Florida, where the *Early Literacy and Learning Model (ELLM)* was developed. These researchers were selected to receive funding for their PCER research projects in a competitive grant application process. Each research team implemented its curriculum and conducted site-specific analyses examining the effects of these curricula on child outcomes. RTI and MPR, the evaluation study contractors, conducted independent evaluations of these and the other treatment curricula that were included in the PCER study. The developers/implementers of these curricula did not conduct the impact analyses that are summarized in this report. Members of the RTI data analysis team completed the impact analyses.

In addition to their role as developers and implementers, Drs. Starkey, Klein, Clements, Sarama, and Lonigan developed measures that were included in the PCER child assessment battery. Drs. Starkey and Klein developed a preschool mathematics assessment, the Child Math Assessment (CMA) that was adapted for use in the PCER evaluation study. The Child Math Assessment-Abbreviated (CMA-A) was added to the assessment battery as a measure of children's early mathematical knowledge and skills using manipulative materials. The *Building Blocks*, Shape Composition task was also included in the child assessment battery. This task was adapted from the *Building Blocks* assessment tool, which was developed by Clements, Sarama, and Liu. The Elision subtest from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) was used in the pre-kindergarten year of the evaluation study. Dr. Christopher Lonigan and his colleagues developed the Pre-CTOPPP, Elision subtest. The assessment was not commercially available at the time it was selected for inclusion in the study or during the data collection phase of the study. A revised version of the assessment became commercially available as the Test of Preschool Early Literacy (TOPEL) in January 2007, after the PCER data collection. Dr. Lonigan has a financial interest in the commercial version of this measure.

Dr. Susan Landry and her colleagues at the Center for Improving the Readiness of Children for Learning and Education (CIRCLE) developed one of the study's classroom observation measures and advised on the selection of the child assessments. CIRCLE staff also trained PCER data collection teams to collect classroom observation data using the Teacher Behavior Rating Scale (TBRS), but CIRCLE staff did not collect the data. CIRCLE staff scored the classroom observation data that were collected using the TBRS measure.

Data collection teams from MPR and RTI independently collected all of the data using the measures that are mentioned here. The data analysis team completed descriptive and impact analyses using the scored data. The developers of these measures had no direct role in the completion of the descriptive analyses or the impact analyses that are summarized in this report.

Page intentionally left blank.

Glossary

ACF—Administration for Children and Families

ANCOVA—Analysis of Covariance

Arnett—Arnett Caregiver Interaction Scale

CMA-A—Child Math Assessment-Abbreviated

control classrooms—Classrooms randomly assigned to the control condition. Classrooms where the prevailing or existing curriculum was in use during the course of the study

control curriculum—The prevailing/existing curriculum used by teachers in the control condition at each site

CTOPP—Comprehensive Test of Phonological Processing (CTOPP), Elision subtest

ECERS-R—Early Childhood Environment Rating Scale-Revised

ECLS-K—Early Childhood Longitudinal Study, Kindergarten cohort

ELLM—*Early Literacy and Learning Model*

FACES—Family and Child Experiences Survey

FSU—Florida State University

full-day—Preschool program where children spend at least 6 hours per day in the preschool classroom

GED—General Educational Development

grantee—Researcher funded by the Institute of Education Sciences, U.S. Department of Education, to conduct a site-specific study under the Preschool Curriculum Evaluation Research initiative. Grants were awarded to investigators at a single institution or to co-investigators at multiple institutions

half-day—Preschool program where children spend less than 6 hours per day

Head Start center—Preschool that is funded by the U.S. Administration for Children and Families Head Start Bureau

ICC—Intraclass correlation

IES—Institute of Education Sciences, U.S. Department of Education

LBS—Learning Behaviors Scale

MDE—Minimum Detectable Effects

MPR—Mathematica Policy Research, Inc

Glossary—Continued

MPR evaluation sites—Preschool Curriculum Evaluation Research research sites where Mathematica Policy Research, Inc. conducted data collection

PCER—Preschool Curriculum Evaluation Research

PLBS—Preschool Learning Behaviors Scale

PPVT—Peabody Picture Vocabulary Test, Third Edition (PPVT-III)

Pre-CTOPPP—Preschool Comprehensive Test of Phonological and Print Processing, Elision subtest

private pre-kindergarten—Preschool that is funded primarily through tuition or other nongovernmental source

public pre-kindergarten—Preschool that is part of a public school system or receives substantial public funding

random assignment—Determination by lottery under supervision of a researcher whether a study subject will be placed in one experimental group or another

randomized trial—Research study in which subjects are randomly assigned to receive or not receive interventions

research site—Collection of preschool programs/classrooms in a specific geographic location that were recruited by each grantee. Grantees implemented one or more preschool curricula at each research site

RTI—RTI International

RTI evaluation sites—PCER research sites where RTI International conducted data collection

SFA—*Success for All*

site/grantee site—The geographic location of the research sites

SSRS—Social Skills Rating System

SSRS Problem Behaviors—Social Skills Rating System, Problem Behaviors scale

SSRS Social Skills—Social Skills Rating System, Social Skills scale

TBRs—Teacher Behavior Rating Scale

TERA—Test of Early Reading Ability, Third Edition (TERA-3)

TOLD—Test of Language Development-Primary, Third Edition (TOLD-P:3)

treatment classroom—Classrooms randomly assigned to the treatment condition where an experimental curriculum was implemented and evaluated

Glossary—Continued

treatment curriculum—One of the 14 intervention curricula that were implemented in treatment classrooms

UNF—University of North Florida

WJ—Woodcock Johnson Achievement Test, 3rd Edition (WJ III)

WJ Applied Problems/WJ Applied Problems test—Woodcock Johnson Achievement Test, 3rd Edition (WJ III), Applied Problems Test

WJ Letter Word Identification/WJ Word Identification test—Woodcock Johnson Achievement Test, 3rd Edition (WJ III), Letter Word Identification Test

WJ Spelling/WJ Spelling test—Woodcock Johnson Achievement Test, 3rd Edition (WJ III), Spelling Test

Page intentionally left blank.

Contents

	Page
Preschool Curriculum Evaluation Research Consortium	iii
Acknowledgments	v
Disclosure of Potential Conflicts of Interest	vii
Glossary	ix
List of Tables	xxi
List of Figures	xxix
Executive Summary	xxxi
Research Questions	xxxi
Study Design	xxxii
Sample and Assignment to Condition	xxxii
Measures	xxxv
Study Implementation	xxxviii
Analysis	xxxix
Results	xli
Chapter 1. An Overview of the Preschool Curriculum Evaluation Research Initiative	1
Study Background	1
School Readiness and Later Academic Achievement	1
Early Childhood Education	2
The Preschool Curriculum Evaluation Research Initiative	3
Research Questions.....	3
Study Design	4
Intervention and Control Curricula	4
Sample and Random Assignment to Condition	6
Measures	9
Study Implementation	16
Timeline of Implementation	17
Response Rates, Attrition, and Mobility	23
Contamination	24
Fidelity of Implementation	25
Sample Description	25
Analysis	26
Results	30
Model Results	31
Considerations: Efficacy, Power, and Multiple Comparisons	31
Criteria for Findings.....	32
Findings	33
Findings by Outcome	33
Findings by Curriculum	38

Contents—Continued

	Page
Chapter 2. <i>Bright Beginnings</i> and <i>Creative Curriculum</i> : Vanderbilt University (Tennessee site)	41
Curriculum	41
<i>Bright Beginnings</i>	41
<i>Creative Curriculum</i>	41
Sample	42
Children and Families	42
Teachers	44
Programs/Classrooms	44
Random Assignment	44
Contamination	45
Control Condition	45
Data Collection	45
Attrition	45
Implementation	46
Implementation Fidelity Ratings	46
Impact Analysis Results	46
<i>Bright Beginnings</i> —Child Outcomes	47
<i>Bright Beginnings</i> —Classroom Outcomes	48
Summary of Findings for <i>Bright Beginnings</i>	49
<i>Creative Curriculum</i> —Child Outcomes	51
<i>Creative Curriculum</i> —Classroom Outcomes	52
Summary of Findings for <i>Creative Curriculum</i>	53
Chapter 3. <i>Creative Curriculum</i> : University of North Carolina at Charlotte (North Carolina and Georgia sites)	55
Curriculum	55
Sample	55
Children and Families	55
Teachers	57
Programs/Classrooms	57
Random Assignment	58
Contamination	59
Control Condition	59
Data Collection	59
Attrition	59
Implementation	59
Implementation Fidelity Ratings	60
Impact Analysis Results	60
<i>Creative Curriculum</i> —Child Outcomes	60
<i>Creative Curriculum</i> —Classroom Outcomes	62
Summary of Findings for <i>Creative Curriculum</i> (North Carolina and Georgia)	63

Contents—Continued

	Page
Chapter 4. <i>Creative Curriculum with Ladders to Literacy</i> : University of New Hampshire (New Hampshire site)	65
Curriculum	65
Sample	65
Children and Families	65
Teachers	67
Programs/Classrooms	67
Random Assignment	67
Contamination	68
Control Condition	68
Data Collection	68
Attrition	69
Implementation	69
Implementation Fidelity Ratings	69
Impact Analysis Results	70
<i>Creative Curriculum with Ladders to Literacy</i> —Child Outcomes	70
<i>Creative Curriculum with Ladders to Literacy</i> —Classroom Outcomes	71
Summary of Findings for <i>Creative Curriculum with Ladders to Literacy</i>	72
Chapter 5. <i>Curiosity Corner: Success for All Foundation</i> (SFA sites: Florida, Kansas, and New Jersey)	75
Curriculum	75
Sample	75
Children and Families	75
Teachers	76
Programs/Classrooms	76
Random Assignment	76
Contamination	78
Control Condition	78
Data Collection	78
Attrition	79
Implementation	79
Implementation Fidelity Ratings	79
Impact Analysis Results	79
<i>Curiosity Corner</i> —Child Outcomes	79
<i>Curiosity Corner</i> —Classroom Outcomes	81
Summary of Findings for <i>Curiosity Corner</i>	82

Contents—Continued

	Page
Chapter 6. <i>Doors to Discovery</i> and <i>Let’s Begin with the Letter People</i> : University of Texas Health Science Center at Houston (Texas site)	85
Curriculum	85
<i>Doors to Discovery</i>	85
<i>Let’s Begin with the Letter People</i>	85
Sample	85
Children and Families	86
Teachers	87
Programs/Classrooms	87
Random Assignment	88
Contamination	89
Control Condition	89
Data Collection	89
Attrition	89
Implementation	90
Implementation Fidelity Ratings	90
Impact Analysis Results	90
<i>Doors to Discovery</i> —Child Outcomes	90
<i>Doors to Discovery</i> —Classroom Outcomes	92
Summary of Findings for <i>Doors to Discovery</i>	93
<i>Let’s Begin with the Letter People</i> —Child Outcomes	95
<i>Let’s Begin with the Letter People</i> —Classroom Outcomes	96
Summary of Findings for <i>Let’s Begin with the Letter People</i>	97
Chapter 7. <i>Early Literacy and Learning Model (ELLM)</i> : University of North Florida (Florida-UNF site)	99
Curriculum	99
Sample	99
Children and Families	100
Teachers	100
Programs/Classrooms	101
Random Assignment	101
Contamination	103
Control Condition	103
Data Collection	103
Attrition	103
Implementation	104
Site-Specific Fidelity Ratings	104
Implementation Fidelity Ratings	104
Impact Analysis Results	105
<i>Early Literacy and Learning Model</i> —Child Outcomes	105
<i>Early Literacy and Learning Model</i> —Classroom Outcomes	106
Summary of Findings for <i>Early Literacy and Learning Model</i>	107

Contents—Continued

	Page
Chapter 8. <i>Language-Focused Curriculum: University of Virginia (Virginia site)</i>	109
Curriculum	109
Sample	109
Children and Families	109
Teachers	110
Programs/Classrooms	110
Random Assignment	110
Contamination	112
Control Condition	112
Data Collection	112
Attrition	112
Implementation	113
Implementation Fidelity Ratings	113
Impact Analysis Results	113
<i>Language-Focused Curriculum</i> —Child Outcomes	113
<i>Language-Focused Curriculum</i> —Classroom Outcomes	115
Summary of Findings for <i>Language-Focused Curriculum</i>	115
Chapter 9. <i>Literacy Express and DLM Early Childhood Express supplemented with Open Court Reading Pre-K: Florida State University (Florida-FSU site)</i>	117
Curriculum	117
<i>Literacy Express</i>	117
<i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i>	117
Sample	118
Children and Families	118
Teachers	118
Programs/Classrooms	118
Random Assignment	120
Contamination	121
Control Condition	121
Data Collection	121
Attrition	121
Implementation	121
Implementation Fidelity Ratings	122
Impact Analysis Results	123
<i>Literacy Express</i> —Child Outcomes	123
<i>Literacy Express</i> —Classroom Outcomes	124
Summary of Findings for <i>Literacy Express</i>	125
<i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> —Child Outcomes	127
<i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> —Classroom Outcomes	128
Summary of Findings for <i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i>	130

Contents—Continued

	Page
Chapter 10. <i>Pre-K Mathematics supplemented with DLM Early Childhood Express</i> <i>Math software: University of California, Berkeley/University at Buffalo,</i> <i>State University of New York (California/New York sites)</i>	131
Curriculum	131
Sample	131
Children and Families	132
Teachers	132
Programs/Classrooms	132
Random Assignment	135
Contamination	136
Control Condition	136
Data Collection	136
Attrition	137
Implementation	137
Implementation Fidelity Ratings	138
Impact Analysis Results	138
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i> —Child Outcomes	138
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i> —Classroom Outcomes	140
Summary of Findings for <i>Pre-K Mathematics supplemented with DLM Early</i> <i>Childhood Express Math software</i>	141
Chapter 11. <i>Project Approach: Purdue University/University of Wisconsin</i> <i>(Wisconsin site)</i>	143
Curriculum	143
Sample	143
Children and Families	143
Teachers	145
Programs/Classrooms	145
Random Assignment	146
Contamination	146
Control Condition	146
Data Collection	146
Attrition	146
Implementation	146
Implementation Fidelity Ratings	148
Impact Analysis Results	148
<i>Project Approach</i> —Child Outcomes	148
<i>Project Approach</i> —Classroom Outcomes	149
Summary of Findings for <i>Project Approach</i>	150

Contents—Continued

	Page
Chapter 12. <i>Project Construct</i> : University of Missouri-Columbia (Missouri site)	153
Curriculum	153
Sample	153
Children and Families	153
Teachers	154
Programs/Classrooms	154
Random Assignment	154
Contamination	156
Control Condition	156
Data Collection	156
Attrition	156
Implementation	157
Implementation Fidelity Ratings	157
Impact Analysis Results	158
<i>Project Construct</i> —Child Outcomes	158
<i>Project Construct</i> —Classroom Outcomes	159
Summary of Findings for <i>Project Construct</i>	160
Chapter 13. <i>Ready, Set, Leap!</i> University of California, Berkeley (New Jersey site)	163
Curriculum	163
Sample	163
Children and Families	164
Teachers	165
Programs/Classrooms	165
Random Assignment	166
Contamination	167
Control Condition	167
Data Collection	167
Attrition	167
Implementation	167
Implementation Fidelity Ratings	167
Impact Analysis Results	167
<i>Ready, Set, Leap!</i> —Child Outcomes	168
<i>Ready, Set, Leap!</i> —Classroom Outcomes	169
Summary of Findings for <i>Ready, Set, Leap!</i>	170
References	173
Appendix A: Secondary Analyses Results	A-1
Appendix B: Data Analysis Approach and Statistical Model	B-1
Appendix C: Unadjusted Mean Scores	C-1
Appendix D: Covariate Adjusted Mean Differences and Standard Errors	D-1

Page intentionally left blank.

List of Tables

Table	Page
Executive Summary	
A The intervention curricula	xxxiii
B Units of random assignment for evaluation of each curriculum	xxxiv
C Outcomes and measures	xxxvi
D Model used with each measure	xl
E Effect sizes for student-level measures: Pre-kindergarten	xlii
F Effect sizes for student-level measures: Kindergarten	xliii
G Effect sizes for classroom-level measures: Pre-kindergarten	xliv
H Findings by student-level outcomes	xlvii
I Findings by classroom-level outcomes	xlviii
Chapter 1. An Overview of the Preschool Curriculum Evaluation Research Initiative	
1.1 The intervention curricula	5
1.2 The intervention and control curricula	6
1.3 Units of random assignment for evaluation of each curriculum	7
1.4 Dispersion of the preschool study sample into kindergarten schools and classrooms	9
1.5 Outcomes and measures	11
1.6 Standardized mean and reliability for outcome measures	12
1.7 Training and support of treatment teachers	20
1.8 Inter-pair agreement on classroom observations among research teams working with RTI International (RTI), fall 2003 and spring 2004	22
1.9 Response rates	23
1.10 Characteristics of children and parents	26
1.11 Characteristics of preschool and kindergarten teachers	27
1.12 Characteristics of preschools	27
1.13 Characteristics of kindergartens	28
1.14 Model used with each outcome measure	29
1.15 Effect sizes for student-level measures: Pre-kindergarten	34
1.16 Effect sizes for student-level measures: Kindergarten	35
1.17 Effect sizes for classroom-level measures: Pre-kindergarten	36
1.18 Achieved minimum detectable effects on the reading, language, mathematics, and behavior composites of measures	37
1.19 Findings by student-level outcomes	37
1.20 Findings by classroom-level outcomes, pre-kindergarten year only	38
Chapter 2. <i>Bright Beginnings</i> and <i>Creative Curriculum</i>: Vanderbilt University (Tennessee site)	
2.1 Child demographic characteristics for <i>Bright Beginnings</i> and <i>Creative Curriculum</i>	42
2.2 Primary caregiver demographic characteristics for <i>Bright Beginnings</i> and <i>Creative Curriculum</i>	43
2.3 Preschool teacher characteristics for <i>Bright Beginnings</i> and <i>Creative Curriculum</i>	44
2.4 Effect sizes for <i>Bright Beginnings</i>	50
2.5 Effect sizes for <i>Creative Curriculum</i> : Tennessee	54

List of Tables—Continued

Table	Page
Chapter 3. <i>Creative Curriculum</i>: University of North Carolina at Charlotte (North Carolina and Georgia sites)	
3.1 Child demographic characteristics for <i>Creative Curriculum</i> : North Carolina and Georgia.....	56
3.2 Primary caregiver demographic characteristics for <i>Creative Curriculum</i> : North Carolina and Georgia	57
3.3 Preschool teacher characteristics for <i>Creative Curriculum</i> : North Carolina and Georgia	58
3.4 Effect sizes for <i>Creative Curriculum</i> : North Carolina and Georgia	64
Chapter 4. <i>Creative Curriculum with Ladders to Literacy</i>: University of New Hampshire (New Hampshire site)	
4.1 Child demographic characteristics for <i>Creative Curriculum with Ladders to Literacy</i>	66
4.2 Primary caregiver demographic characteristics for <i>Creative Curriculum with Ladders to Literacy</i>	66
4.3 Preschool teacher characteristics for <i>Creative Curriculum with Ladders to Literacy</i>	68
4.4 Effect sizes for <i>Creative Curriculum with Ladders to Literacy</i>	73
Chapter 5. <i>Curiosity Corner: Success for All Foundation</i> (SFA sites: Florida, Kansas, and New Jersey)	
5.1 Child demographic characteristics for <i>Curiosity Corner</i>	76
5.2 Primary caregiver demographic characteristics for <i>Curiosity Corner</i>	77
5.3 Preschool teacher characteristics for <i>Curiosity Corner</i>	78
5.4 Effect sizes for <i>Curiosity Corner</i>	83
Chapter 6. <i>Doors to Discovery</i> and <i>Let's Begin with the Letter People</i>: University of Texas Health Science Center at Houston (Texas site)	
6.1 Child demographic characteristics for <i>Doors to Discovery</i> and <i>Let's Begin with the Letter People</i>	86
6.2 Primary caregiver demographic characteristics for <i>Doors to Discovery</i> and <i>Let's Begin with the Letter People</i>	87
6.3 Preschool teacher characteristics for <i>Doors to Discovery</i> and <i>Let's Begin with the Letter People</i>	88
6.4 Effect sizes for <i>Doors to Discovery</i>	94
6.5 Effect sizes for <i>Let's Begin with the Letter People</i>	98
Chapter 7. <i>Early Literacy and Learning Model (ELLM)</i>: University of North Florida (Florida-UNF site)	
7.1 Child demographic characteristics for <i>Early Literacy and Learning Model</i>	100
7.2 Primary caregiver demographic characteristics for <i>Early Literacy and Learning Model</i>	101
7.3 Preschool teacher characteristics for <i>Early Literacy and Learning Model</i>	102
7.4 Effect sizes for <i>Early Literacy and Learning Model</i>	108
Chapter 8. <i>Language-Focused Curriculum</i>: University of Virginia (Virginia site)	
8.1 Child demographic characteristics for <i>Language-Focused Curriculum</i>	110
8.2 Primary caregiver demographic characteristics for <i>Language-Focused Curriculum</i>	111
8.3 Preschool teacher characteristics for <i>Language-Focused Curriculum</i>	112
8.4 Effect sizes for <i>Language-Focused Curriculum</i>	116

List of Tables—Continued

Table	Page
Chapter 9. <i>Literacy Express</i> and <i>DLM Early Childhood Express</i> supplemented with <i>Open Court Reading Pre-K</i>: Florida State University (Florida-FSU site)	
9.1 Child demographic characteristics for <i>Literacy Express</i> and <i>DLM Early Childhood Express</i> supplemented with <i>Open Court Reading Pre-K</i>	119
9.2 Primary caregiver demographic characteristics for <i>Literacy Express</i> and <i>DLM Early Childhood Express</i> supplemented with <i>Open Court Reading Pre-K</i>	119
9.3 Preschool teacher characteristics for <i>Literacy Express</i> and <i>DLM Early Childhood Express</i> supplemented with <i>Open Court Reading Pre-K</i>	120
9.4 Effect sizes for <i>Literacy Express</i>	126
9.5 Effect sizes for <i>DLM Early Childhood Express</i> supplemented with <i>Open Court Reading Pre-K</i>	130
Chapter 10. <i>Pre-K Mathematics</i> supplemented with <i>DLM Early Childhood Express Math software</i>: University of California, Berkeley/University at Buffalo, State University of New York (California/New York sites)	
10.1 Child demographic characteristics for <i>Pre-K Mathematics</i> supplemented with <i>DLM Early Childhood Express Math software</i>	133
10.2 Primary caregiver demographic characteristics for <i>Pre-K Mathematics</i> supplemented with <i>DLM Early Childhood Express Math software</i>	134
10.3 Preschool teacher characteristics for <i>Pre-K Mathematics</i> supplemented with <i>DLM Early Childhood Express Math software</i>	135
10.4 Effect sizes for <i>Pre-K Mathematics</i> supplemented with <i>DLM Early Childhood Express Math software</i>	142
Chapter 11. <i>Project Approach</i>: Purdue University/University of Wisconsin (Wisconsin site)	
11.1 Child demographic characteristics for <i>Project Approach</i>	144
11.2 Primary caregiver demographic characteristics for <i>Project Approach</i>	144
11.3 Preschool teacher characteristics for <i>Project Approach</i>	145
11.4 Effect sizes for <i>Project Approach</i>	151
Chapter 12. <i>Project Construct</i>: University of Missouri-Columbia (Missouri site)	
12.1 Child demographic characteristics for <i>Project Construct</i>	154
12.2 Primary caregiver demographic characteristics for <i>Project Construct</i>	155
12.3 Preschool teacher characteristics for <i>Project Construct</i>	156
12.4 Effect sizes for <i>Project Construct</i>	161
Chapter 13. <i>Ready, Set, Leap!</i> University of California, Berkeley (New Jersey site)	
13.1 Child demographic characteristics for <i>Ready, Set, Leap!</i>	164
13.2 Primary caregiver demographic characteristics for <i>Ready, Set, Leap!</i>	165
13.3 Preschool teacher characteristics for <i>Ready, Set, Leap!</i>	166
13.4 Effect sizes for <i>Ready, Set, Leap!</i>	171
Appendix A. Secondary Analyses Results	
A-1 Possible early treatment effects and non-equivalence at baseline	A-4
A-2 Secondary analysis: Outcomes, measures, models, and grades analyzed	A-5
A-3 Correlation matrix for student-level measures	A-8
A-4 Criterion used to determine curricula’s impact on a measure and on an outcome	A-9

List of Tables—Continued

Table	Page
Appendix A. Secondary Analyses Results—Continued	
A-5 Findings on student-level outcomes: Main and secondary analyses	A-11
A-6 Findings on classroom-level outcomes: Main and secondary analyses	A-12
A-7 Secondary analysis results for <i>Bright Beginnings</i>	A-18
A-8 Secondary analysis results for <i>Creative Curriculum: Tennessee</i>	A-24
A-9 Secondary analysis results for <i>Creative Curriculum: North Carolina and Georgia</i>	A-30
A-10 Secondary analysis results for <i>Creative Curriculum with Ladders to Literacy</i>	A-36
A-11 Secondary analysis results for <i>Curiosity Corner</i>	A-42
A-12 Secondary analysis results for <i>Doors to Discovery</i>	A-48
A-13 Secondary analysis results for <i>Let's Begin with the Letter People</i>	A-54
A-14 Secondary analysis results for <i>Early Literacy and Learning Model</i>	A-60
A-15 Secondary analysis results for <i>Language-Focused Curriculum</i>	A-65
A-16 Secondary analysis results for <i>Literacy Express</i>	A-71
A-17 Secondary analysis results for <i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i>	A-78
A-18 Secondary analysis results for <i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i>	A-84
A-19 Secondary analysis results for <i>Project Approach</i>	A-90
A-20 Secondary analysis results for <i>Project Construct</i>	A-96
A-21 Secondary analysis results for <i>Ready, Set, Leap!</i>	A-102
Appendix B. Data Analysis Approach and Statistical Model	
B-1 Units of random assignment for evaluation of each curriculum	B-4
B-2 Variables in analysis	B-6
B-3 Response rates and attrition	B-9
B-4 Main analysis: Model used with each outcome measure	B-12
B-5 Secondary analysis: Models and grades for each outcome measure	B-13
B-6 Average estimated population standard deviation by study outcome and time points	B-28
B-7 Average estimated population standard deviation by study outcome and group assignment	B-29
B-8 Pooled standard deviation example	B-29
B-9 Estimated pooled population standard deviation using unconditional standard deviations and standard deviations from repeated measures analyses	B-31
B-10 Pooled standard deviation details: Pooling for outcomes modeled with the simple repeated measures, the repeated measures spline models, the pre-kindergarten spring analysis of covariance (ANCOVA) models (except Teacher Behavior Rating Scale [TBRS]), and the kindergarten spring ANCOVA models where kindergarten data were comparable to pre-kindergarten	B-32
B-11 Pooled standard deviation details: Pooling for Teacher Behavior Rating Scale (TBRS) outcomes modeled with the pre-kindergarten spring analysis of covariance (ANCOVA) models	B-33
B-12 Pooled standard deviation details: Pooling for kindergarten spring outcomes (SSRS, Pre-CTOPPP/CTOPP, PLBS/LBS) modeled with analysis of covariance (ANCOVA) models	B-34
B-13 Pre-kindergarten and kindergarten classroom clusters of children, maximum and minimum size	B-36
B-14 Specific covariance structures found to best fit the data	B-38

List of Tables—Continued

Table	Page
Appendix B. Data Analysis Approach and Statistical Model—Continued	
B-15 Specific covariance structure definitions	B-38
B-16 Significant treatment by covariate interactions from check on homogeneity of regression assumption.....	B-40
 Appendix C. Unadjusted Mean Scores	
C-1a Unadjusted mean scores of child-level outcome measures, <i>Bright Beginnings</i> : Tennessee	C-3
C-1b Unadjusted mean scores of classroom-level outcome measures, <i>Bright Beginnings</i> : Tennessee	C-4
C-2a Unadjusted mean scores of child-level outcome measures, <i>Creative Curriculum</i> : Tennessee	C-5
C-2b Unadjusted mean scores of classroom-level outcome measures, <i>Creative Curriculum</i> : Tennessee	C-6
C-3a Unadjusted mean scores of child-level outcome measures, <i>Creative Curriculum</i> : North Carolina and Georgia	C-7
C-3b Unadjusted mean scores of classroom-level outcome measures, <i>Creative Curriculum</i> : North Carolina and Georgia	C-8
C-4a Unadjusted mean scores of child-level outcome measures, <i>Creative Curriculum with Ladders to Literacy</i> : New Hampshire	C-9
C-4b Unadjusted mean scores of classroom-level outcome measures, <i>Creative Curriculum with Ladders to Literacy</i> : New Hampshire	C-10
C-5a Unadjusted mean scores of child-level outcome measures, <i>Curiosity Corner</i> : Florida, Kansas, and New Jersey	C-11
C-5b Unadjusted mean scores of classroom-level outcome measures, <i>Curiosity Corner</i> : Florida, Kansas, and New Jersey	C-12
C-6a Unadjusted mean scores of child-level outcome measures, <i>Doors to Discovery</i> : Texas	C-13
C-6b Unadjusted mean scores of classroom-level outcome measures, <i>Doors to Discovery</i> : Texas	C-14
C-7a Unadjusted mean scores of child-level outcome measures, <i>Let's Begin with the Letter People</i> : Texas	C-15
C-7b Unadjusted mean scores of classroom-level outcome measures, <i>Let's Begin with the Letter People</i> : Texas	C-16
C-8a Unadjusted mean scores of child-level outcome measures, <i>Early Literacy and Learning Model</i> : Florida—University of North Florida	C-17
C-8b Unadjusted mean scores of classroom-level outcome measures, <i>Early Literacy and Learning Model</i> : Florida—University of North Florida	C-18
C-9a Unadjusted mean scores of child-level outcome measures, <i>Language-Focused Curriculum</i> : Virginia	C-19
C-9b Unadjusted mean scores of classroom-level outcome measures, <i>Language-Focused Curriculum</i> : Virginia	C-20
C-10a Unadjusted mean scores of child-level outcome measures, <i>Literacy Express</i> : Florida—Florida State University	C-21
C-10b Unadjusted mean scores of classroom-level outcome measures, <i>Literacy Express</i> : Florida—Florida State University	C-22
C-11a Unadjusted mean scores of child-level outcome measures, <i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> : Florida—Florida State University	C-23
C-11b Unadjusted mean scores of classroom-level outcome measures, <i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> : Florida—Florida State University	C-24

List of Tables—Continued

Table	Page
Appendix C. Unadjusted Mean Scores—Continued	
C-12a	Unadjusted mean scores of child-level outcome measures, <i>Pre-K Mathematics supplemented with DLM Express Math software</i> : California and New York C-25
C-12b	Unadjusted mean scores of classroom-level outcome measures, <i>Pre-K Mathematics supplemented with DLM Express Math software</i> : California and New York C-26
C-13a	Unadjusted mean scores of child-level outcome measures, <i>Project Approach</i> : Wisconsin C-27
C-13b	Unadjusted mean scores of classroom-level outcome measures, <i>Project Approach</i> : Wisconsin C-28
C-14a	Unadjusted mean scores of child-level outcome measures, <i>Project Construct</i> : Missouri C-29
C-14b	Unadjusted mean scores of classroom-level outcome measures, <i>Project Construct</i> : Missouri C-30
C-15a	Unadjusted mean scores of child-level outcome measures, <i>Ready, Set, Leap!</i> : New Jersey C-31
C-15b	Unadjusted mean scores of classroom-level outcome measures, <i>Ready, Set, Leap!</i> : New Jersey C-32
Appendix D. Covariate Adjusted Mean Differences and Standard Errors	
D-1a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Bright Beginnings</i> : Tennessee D-3
D-1b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Bright Beginnings</i> : Tennessee D-4
D-2a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Creative Curriculum</i> : Tennessee D-5
D-2b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Creative Curriculum</i> : Tennessee D-6
D-3a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Creative Curriculum</i> : North Carolina and Georgia D-7
D-3b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Creative Curriculum</i> : North Carolina and Georgia D-8
D-4a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Creative Curriculum with Ladders to Literacy</i> : New Hampshire D-9
D-4b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Creative Curriculum with Ladders to Literacy</i> : New Hampshire D-10
D-5a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Curiosity Corner</i> : Florida, Kansas, and New Jersey D-11
D-5b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Curiosity Corner</i> : Florida, Kansas, and New Jersey D-12
D-6a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Doors to Discovery</i> : Texas D-13
D-6b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Doors to Discovery</i> : Texas D-14
D-7a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Let's Begin with the Letter People</i> : Texas D-15
D-7b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Let's Begin with the Letter People</i> : Texas D-16
D-8a	Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Early Literacy and Learning Model</i> : Florida—University of North Florida D-17
D-8b	Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Early Literacy and Learning Model</i> : Florida—University of North Florida D-18

List of Tables—Continued

Table	Page
Appendix D. Covariate Adjusted Mean Differences and Standard Errors—Continued	
D-9a Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Language-Focused Curriculum</i> : Virginia	D-19
D-9b Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Language-Focused Curriculum</i> : Virginia	D-20
D-10a Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Literacy Express</i> : Florida—Florida State University	D-21
D-10b Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Literacy Express</i> : Florida—Florida State University	D-22
D-11a Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> : Florida—Florida State University	D-23
D-11b Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> : Florida—Florida State University	D-24
D-12a Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i> : California and New York	D-25
D-12b Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i> : California and New York	D-26
D-13a Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Project Approach</i> : Wisconsin	D-27
D-13b Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Project Approach</i> : Wisconsin	D-28
D-14a Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Project Construct</i> : Missouri	D-29
D-14b Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Project Construct</i> : Missouri	D-30
D-15a Covariate adjusted mean differences and standard errors of child-level outcome measures, <i>Ready, Set, Leap!</i> : New Jersey	D-31
D-15b Covariate adjusted mean differences and standard errors of classroom-level outcome measures, <i>Ready, Set, Leap!</i> : New Jersey	D-32

Page intentionally left blank.

List of Figures

Figure		Page
1.1	Timeline for teams working with RTI International	18
1.2	Timeline for teams working with Mathematica Policy Research, Inc.	19
B-1	Repeated measures spline model	B-15
B-2	Simple repeated measures model	B-15
B-3	Pre-kindergarten (Pre-K) and kindergarten (K) analysis of covariance (ANCOVA) models	B-17

Page intentionally left blank.

Executive Summary

A variety of preschool curricula is available and in widespread use, however, there is a lack of evidence from rigorous evaluations regarding the effects of these curricula on children's school readiness. The lack of such information is important as early childhood center-based programs have been a major, sometimes the sole, component of a number of federal and state efforts to improve young at-risk children's school readiness (e.g., Head Start, Even Start, public pre-kindergarten). In 2005, nearly half (47%) of all 3- to 5-year-old children from low-income families were enrolled in either part-day or full-day early childhood programs (U.S. Department of Education 2006).

In 2002, the Institute of Education Sciences (IES) began the Preschool Curriculum Evaluation Research (PCER) initiative to conduct rigorous efficacy evaluations of available preschool curricula. Twelve research teams implemented one or two curricula in preschool settings serving predominantly low-income children under an experimental design. For each team, preschools or classrooms were randomly assigned to the intervention curricula or control curricula and the children were followed from pre-kindergarten through kindergarten. IES contracted with RTI International (RTI) and Mathematica Policy Research (MPR) to evaluate the impact of each of the 14 curricula implemented using a common set of measures with the cohort of children beginning preschool in the summer-fall of 2003.

This report provides the individual results for each curriculum from the evaluations by RTI and MPR. Chapter 1 describes the PCER initiative and details the common elements of the evaluations including the experimental design, implementation, analysis, results, and findings. Chapters 2-13, respectively, provide greater detail on the individual evaluations of the curricula implemented by each research team including information on the curricula, the demographics of the site-specific samples, assignment, fidelity of implementation, and results. Appendix A presents results from a secondary analysis of the data. Appendix B provides greater detail regarding the data analyses conducted. Appendixes C and D provide additional information regarding the outcome measures.

Research Questions

The PCER initiative focused on the impact of the intervention curricula on students' reading and pre-reading, phonological awareness, early language, early mathematics knowledge, and behavior (including social skills and problem behaviors) at the end of pre-kindergarten and kindergarten. These domains of knowledge and skills are predictive of academic success in the early years of elementary school (Downer and Pianta 2006; Miles and Stipek 2006). As a result, the research questions for the initiative primarily concern student outcomes and also include classroom outcomes due to their potentially mediating or moderating roles. The research questions are:

1. What is the impact of each of the 14 preschool curricula on preschool students' early reading skills, phonological awareness, language development, early mathematical knowledge, and behavior?
2. What is the impact of each of the 14 preschool curricula on these outcomes for students at the end of kindergarten?
3. What is the impact of each of the 14 preschool curricula on preschool classroom quality, teacher-child interaction, and instructional practices?

Study Design

Under the PCER initiative, 12 research teams received peer-reviewed grants to implement one to two preschool curricula of their choosing under an experimental design. For each team's evaluation, preschool classrooms or programs were randomly assigned to use the treatment or control curricula. The treatment curricula included sufficient standardized training procedures and curriculum materials to be implemented in typical early childhood education settings. RTI and MPR evaluated the impact of each curriculum using a common set of measures. The curricula, corresponding research team, research site, and evaluator are listed in table A. Three teams each implemented two curricula. Two teams implemented the same curriculum, *Creative Curriculum*. Four teams had originally developed the curricula that they implemented (*Curiosity Corner*, *Literacy Express*, *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*, and *Early Literacy and Learning Model [ELLM]*). RTI evaluated eight curricula implemented by seven teams (including one curriculum that was evaluated by two teams) while MPR evaluated six curricula implemented by five teams. In sum, 14 curricula (one twice) were evaluated.

The 14 curricula were evaluated in comparison to the local control condition that, in general, was the local curriculum-as-usual. As a result, multiple curricula were used across the control sites and within some of the individual evaluations. These included teacher-developed nonspecific curricula with a focus on basic school readiness, district-developed curricula, and published curricula (some of which were implemented by other research teams). The control curricula are identified in the section on Findings by Curriculum at the end of the Executive Summary. As a result of the use of different control curricula among the evaluations, this report does not make cross-intervention comparisons.

Rather than one overall evaluation, the PCER study contains individual evaluations for each curriculum, for three reasons. First, each research team worked independently. Second, the selection of the intervention and the randomized assignment occurred at the team level. Third, different control curricula were used with each intervention curriculum.

Sample and Assignment to Condition

Preschool programs taking part in the evaluation of the curricula included Head Start centers, private child care centers, and public pre-kindergarten programs in urban, rural, and suburban locations. Each research team recruited interested local preschool programs. IES had set a funding priority on grant applications that addressed preschools serving children from low-income families, with the result that 88 percent of the preschools included were either Head Start centers or public pre-kindergarten programs, and half of the children's primary caregivers had a high school education or less. Programs agreed to the random assignment (by program or classroom) to a treatment curriculum or to local control conditions.

For each evaluated curriculum, table B indicates whether pre-kindergarten programs or classrooms were randomly assigned to treatment or control conditions, the number assigned to each, and the number of treatment and control students included in each evaluation. Three teams (implementing four curricula) randomly assigned pre-kindergarten programs, and the other nine teams randomly assigned classrooms. Three teams compared two curricula against a single set of control classrooms or programs. All but two teams (Purdue University and University of New Hampshire) used block random assignment.

Table A. The intervention curricula

Curriculum and publisher	Research team	Research site	Evaluator
<i>Bright Beginnings</i> (Charlotte-Mecklenburg Schools 2001)	Vanderbilt University	Tennessee	RTI
<i>Creative Curriculum</i> (Teaching Strategies, Inc. 2002)	Vanderbilt University	Tennessee	RTI
<i>Creative Curriculum</i> (Teaching Strategies, Inc. 2002)	University of North Carolina at Charlotte	North Carolina and Georgia	RTI
<i>Creative Curriculum with Ladders to Literacy</i> (Teaching Strategies, Inc. 2002; Paul H. Brookes Publishing Company 1998)	University of New Hampshire	New Hampshire	RTI
<i>Curiosity Corner</i> (Success for All Foundation, Inc. 2003)	Success for All Foundation	Florida, Kansas, New Jersey	MPR
<i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> (SRA/McGraw-Hill 2003)	Florida State University	Florida	MPR
<i>Doors to Discovery</i> (Wright Group/McGraw-Hill 2001)	University of Texas Health Science Center at Houston	Texas	RTI
<i>Early Literacy and Learning Model</i> (Florida Institute of Education and the University of North Florida 2002)	University of North Florida	Florida	RTI
<i>Language-Focused Curriculum</i> (Paul H. Brookes Publishing Company 1995)	University of Virginia	Virginia	MPR
<i>Let's Begin with the Letter People</i> (Abrams & Company 2000)	University of Texas Health Science Center at Houston	Texas	RTI
<i>Literacy Express</i> (Author: Lonigan and Farver 2002, unpublished)	Florida State University	Florida	MPR
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i> (Scott Foresman—Pre-K Mathematics 2002; SRA/ McGraw-Hill—DLM Early Childhood Express Math software 2003)	University of California, Berkeley and University at Buffalo, State University of New York	California and New York	RTI
<i>Project Approach</i> (Ablex 1989)	Purdue University and University of WI-Milwaukee	Wisconsin	RTI
<i>Project Construct</i> (Missouri Department of Elementary and Secondary Education 1992)	University of Missouri- Columbia	Missouri	MPR
<i>Ready, Set, Leap!</i> (LeapFrog School House 2003)	University of California, Berkeley	New Jersey	MPR

NOTE: RTI: RTI International

MPR: Mathematica Policy Research, Inc.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B. Units of random assignment for evaluation of each curriculum

Research team	Curricula	Treatment sample	Control sample	Students
Vanderbilt University	<i>Bright Beginnings</i>	7 classrooms	7 classrooms	T: 103
	<i>Creative Curriculum</i>	7 classrooms		C: 105
University of North Carolina at Charlotte	<i>Creative Curriculum</i>	9 classrooms	9 classrooms	T: 97 C: 97
University of New Hampshire	<i>Creative Curriculum with Ladders to Literacy</i>	7 classrooms	7 classrooms	T: 62 C: 61
Success for All Foundation	<i>Curiosity Corner</i>	10 Pre-K programs	8 Pre-K programs	T: 105 C: 110
University of Texas Health Science Center at Houston	<i>Doors to Discovery</i>	14 classrooms	15 classrooms	T: 101
	<i>Let's Begin with the Letter People</i>	15 classrooms		C: 96
University of North Florida	<i>Early Literacy and Learning Model</i>	14 classrooms ¹	14 classrooms ¹	T: 137 C: 107
University of Virginia	<i>Language-Focused Curriculum</i>	7 classrooms	7 classrooms	T: 97 C: 98
Florida State University	<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>	5 Pre-K programs	6 Pre-K programs	T: 101
	<i>Literacy Express</i>	6 Pre-K programs		C: 97
UC-Berkeley and University at Buffalo, State University of New York	<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>	20 classrooms	20 classrooms	T: 159 C: 157
Purdue University and University of WI-Milwaukee	<i>Project Approach</i>	7 classrooms	6 classrooms	T: 114 C: 90
University of Missouri-Columbia	<i>Project Construct</i>	10 Pre-K programs ¹	11 Pre-K programs ¹	T: 123 C: 108
University of California, Berkeley	<i>Ready, Set, Leap!</i>	18 classrooms	21 classrooms	T: 149 C: 137

¹ After one program or classroom attrited.

NOTE: T: Treatment Group

C: Control Group

Three research teams (Vanderbilt University, University of Texas Health Science Center at Houston, and Florida State University) have two treatment groups and a shared control group. When reading the "Students" column, the first "T" refers to the first curriculum in the same row, while the second "T" refers to the second curriculum in the same row. The "C" refers to the shared control group. For example, Vanderbilt University compared two curricula: *Bright Beginnings* (103 students) and *Creative Curriculum* (101 students) to a control curriculum (105 students).

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

The process of random assignment differed somewhat depending upon the evaluator. The seven research teams working with RTI were responsible for the random assignment at their sites; RTI monitored the process and tracked any changes. These teams had a pilot preschool implementation year starting in the fall of 2002. The randomization conducted in that year carried over, in most cases, to the actual evaluation begun in the 2003-04 school year. The five research teams working with MPR began implementing the curricula in the 2003-04 school year. In conjunction with the research teams, MPR conducted block random assignment for four teams. In addition, Florida State University (FSU) block randomly assigned pre-kindergarten programs to its two curricula and the control group.

The analyses included 2,911 children, 315 preschool classrooms, and 208 preschools. As noted above, the PCER study individually evaluates separate curriculum so no comparisons are made between all those included in the treatment condition and all those who were part of the control condition. Such comparisons are made for each evaluation's treatment and control groups in chapters 2 to 13.

On average, the students were age 4.6 years at the time of the baseline data collection in the fall of 2003 and age 6.1 years at the time of the kindergarten follow-up in the spring of 2005. Approximately half (51%) of the children were male. One-third were white non-Hispanic, 43 percent were African American, and 16 percent were Hispanic. Less than 7 percent had a disability. On average, the students' primary caregivers, most often their biological or adoptive mother, were age 32 years at the time of the fall 2003 data collection. Less than half (47%) were married and one-third were never married. Less than half attended or graduated from college (48%), one-third had a high school diploma or GED, and 19 percent did not complete high school. Half were employed full-time, 14 percent part-time, and 34 percent were unemployed.

Almost all the preschool teachers were female (98%) and the majority were White (54%), with one-third African-American. Two-thirds had at least a college degree. On average, they had 12 years of teaching experience and 8 years of experience teaching in pre-kindergarten settings. A majority (87%) of the preschool programs in which they taught were full-day programs. More than half (58%) were public pre-kindergartens, 31 percent were Head Start teachers, and child care teachers made up the remainder (12%). On average, teachers taught 15 students, with a child-staff ratio averaging 7.5 children per teacher.

The kindergarten teachers were also mostly female (98%) and White (74%), with 17 percent African-American. Almost all had at least a BA (97%) with 39 percent having a graduate degree. They averaged 15 years of teaching experience, with an average of 9 years teaching kindergarten. Ninety-three percent of the kindergarten classrooms were full-day and 92 percent of the students were enrolled in public schools. The average number of students per classroom was 20 children. Thirty-nine percent were enrolled in schools where more than 75 percent of the students were eligible for free or reduced-price lunch.

Measures

Twenty-seven measures were chosen to address the outcomes of interest regarding children's school readiness (reading, phonological awareness, language, mathematics, and behavior) and classroom conditions (classroom quality, teacher-child interaction, and instructional practices). Table C lists the measures used for each outcome, when they were collected, and through which instrument they were collected. Five major data collection instruments were used to collect the outcome measures and other student, school and family data: (1) a child assessment, (2) a teacher report, (3) classroom observation, (4) a teacher interview or questionnaire, and (5) a parent interview.

Child Assessment

The child assessment measured the student-level academic outcomes for the evaluation, beginning with a preschool pre-test in the fall of 2003 and post-tests near the end of preschool in the spring of 2004, and the

Table C. Outcomes and measures

Outcome	Measures	Times collected	Instrument
Reading	TERA WJ Letter Word Identification WJ Spelling	Pre-K: fall/spring, K: spring Pre-K: fall/spring, K: spring Pre-K: fall/spring, K: spring	Child assessment
Pre-kindergarten phonological awareness ¹	Pre-CTOPPP	Pre-K: fall/spring	Child assessment
Kindergarten phonological awareness ¹	CTOPP	K: spring	Child assessment
Language	PPVT TOLD	Pre-K: fall/spring, K: spring Pre-K: fall/spring, K: spring	Child assessment
Mathematics	WJ Applied Problems CMA-A Mathematics Composite Shape Composition ²	Pre-K: fall/spring, K: spring Pre-K: fall/spring, K: spring Pre-K: fall/spring, K: spring	Child assessment
Pre-kindergarten behavior ¹	SSRS Social Skills SSRS Problem Behavior PLBS	Pre-K: fall/spring Pre-K: fall/spring Pre-K: fall/spring	Teacher report
Kindergarten behavior ¹	SSRS Social Skills SSRS Problem Behavior LBS	K: spring K: spring K: spring	Teacher report
Classroom quality	ECERS-R	Pre-K: fall/spring	Classroom observation
Teacher-child interaction	Arnett Detachment Arnett Harshness Arnett Permissiveness Arnett Positive Interaction	Pre-K: fall/spring Pre-K: fall/spring Pre-K: fall/spring Pre-K: fall/spring	Classroom observation
Literacy instruction	TBRS Written Expression TBRS Print and Letter Knowledge	Pre-K: spring Pre-K: spring	Classroom observation
Phonological instruction	TBRS Phonological Awareness	Pre-K: spring	Classroom observation
Language instruction	TBRS Book Reading TBRS Oral Language	Pre-K: spring Pre-K: spring	Classroom observation
Mathematics instruction	TBRS Math Concepts	Pre-K: spring	Classroom observation

¹ Pre-kindergarten and kindergarten measures are not on the same scale.

² Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

end of kindergarten in the spring of 2005. Individually administered, the battery assessed beginning reading skills, phonological awareness, oral language development, and mathematical knowledge and skills. The measures regarding reading included the Test of Early Reading Ability (TERA) (Reid, Hresko, and Hammill 2001), the Woodcock Johnson (WJ) Letter Word Identification, and WJ Spelling (McGrew and Woodcock 2001). For phonological awareness, the measures were the Elision subtests of the Preschool Comprehensive Test of Phonologic and Print Processing and the Comprehensive Test of Phonological Processing for kindergarten (Pre-CTOPPP and CTOPP) (Wagner, Torgeson, and Rashotte 1999). For language, the

measures included the Peabody Picture Vocabulary Test (PPVT) (Dunn and Dunn 1997) and the Test of Language Development (TOLD) Grammatical Understanding subtest (Newcomer and Hammill 1997). For mathematics, the measures were the WJ Applied Problems (McGrew and Woodcock 2001), the Child Math Assessment-Abbreviated (CMA-A) Composite Score (Klein and Starkey 2002), and the Building Blocks' Shape Composition Task (unpublished).

Teacher Report of Child Behavior

Teacher reports provided the student-level behavior measures used in the evaluation. Preschool teachers gave pre-intervention ratings of child behaviors in the fall of 2003 and post-intervention ratings in the spring of 2004. They rated each child's behavior (social competence, behavior problems, and classroom performance) using three scales: the Social Skills Rating System (SSRS) Social Skills scale, the SSRS Problem Behaviors scale (Gresham and Elliott 1990), and the Preschool Learning Behaviors Scale (PLBS) (McDermott et al. 2000). Kindergarten teachers provided a longer-term post-intervention rating on the students' behavior in the spring of 2005 using the two SSRS scales and the Learning Behaviors Scale (LBS) (McDermott et al. 2000).

Classroom Observation

Two pre-intervention classroom measures and three post-intervention classroom measures were gathered from preschool classroom observations. No observations were made of kindergarten classrooms. Three scales designed to characterize the quality and organization of the classroom and the nature of the interaction between children and the teacher were used in the observations. The Early Childhood Environment Rating Scale-Revised (ECERS-R) (Harms, Clifford, and Cryer 1998) provided an overall measure of the quality of the classroom. The Arnett Caregiver Interaction Scale (Arnett) (Arnett 1989) measured teacher-child interaction on four scales: Positive Interaction, Harshness, Detachment, and Permissiveness. The pre-intervention observation using the ECERS-R and Arnett Scale was conducted in the fall of 2003 and the post-intervention observation in the spring of 2004. The Teacher Behavior Rating Scale (TBRS) (Landry et al. 2002) was added as a post-intervention measure to the spring 2004 observation to capture preschool instructional practices. The TBRS includes scales for teacher instructional practices regarding: written expression, print and letter knowledge, phonological awareness, book reading, oral language use, and mathematics concepts.

Teacher Interview/Questionnaire

Preschool teachers were interviewed regarding the types and frequency of classroom activities, general classroom information, clarification of observational data, teacher attitudes and beliefs, and teacher background information. The background information was used to construct covariates for the models used to analyze the data. Instead of an interview, kindergarten teachers completed a questionnaire that addressed their background, views on readiness, classroom resources and activities, instructional practices, and interactions with parents.

Parent Interview

Parents were interviewed regarding demographic information, their own and their child's health and disability status, their assessment of the child's accomplishments and social skills, family-child activities, parenting practices, parental depression, and the use of child care. The interview drew primarily from the Head Start's Family and Child Experiences Survey (FACES) (U.S. Department of Health and Human Services 2002) supplemented with additional measures. The demographic information and disability status were used to construct covariates for the models used to analyze the data.

Study Implementation

The key implementation events in the evaluation of each curriculum included randomization of classrooms or programs, consent gathering, teacher training in the use of a treatment curriculum, implementation of the curriculum in the classroom, training the assessors, and collection of the baseline student and classroom measures and the post-intervention measures in preschool and kindergarten. As research teams independently implemented the curricula and as the schools followed different calendars, the dates and sometimes the order of these events differed between teams and sites within teams.

Randomization for the seven teams working with RTI occurred in the pilot year (starting in the fall of 2002) and mostly carried over into the 2003-04 evaluation year. For the five teams working with MPR, there was no pilot year and their time of randomization ranged from July through September of 2003.

The consent process followed randomization, except for two teams, for which it occurred concurrently. The start of implementation of the curricula in the classroom ranged from August through October 2003. The RTI and MPR data collection teams attempted to collect baseline data close to the beginning of school to avoid student exposure to the treatment curricula before pre-testing. Twelve teams began implementation before baseline data collection and two teams began implementation concurrently with collection. The lag between the start of implementation and the collection of baseline data ranged from 8 to 49 days (appendix A discusses additional analyses to adjust for possible early treatment effects that might result from these cases). Baseline data collection followed the consent process for the teams working with MPR and ran concurrently for the teams working with RTI. Baseline data collection took 6 to 8 weeks between September and November 2003. Assessors were trained the week of August 4, 2003 for the teams working with RTI and the week of September 8, 2003 for the teams working with MPR.

The amount and timing of teacher training varied by team. The teams working with RTI provided most of the training during the 2002 pilot year, then gave refresher training during the 2003 evaluation year. The teams working with MPR provided initial training at the beginning of the evaluation year, and then follow-up training throughout the year. The students' exposure to the treatment curriculum and their teachers' training in its use was confined to preschool for all teams except in the case of the Success for All (SFA) team; in this case, some children entered SFA kindergarten classrooms where the SFA *Kinder Corners* curriculum was in use.

Pre-kindergarten post-test data were collected in the spring, from April to June 2004, depending on school calendars. Student assessments, teacher interviews, teacher reports on behavior, and classroom observations were completed over a 6- to 8-week period. Parent interviews were completed over a 12-week period. Kindergarten post-test data (student assessments, teacher reports, teacher surveys, and parent interviews but no classroom observations) were collected in the spring and summer of 2005 between March and July.

Fidelity of Implementation

The research teams collected data on the fidelity of implementation for the treatment and control curricula using both a team-specific measure and a global implementation rating that can be used for between-curricula comparisons. The global ratings use a four-point scale representing High, Medium, Low, or No Implementation. The fidelity of implementation for both the treatment and control curricula was rated as Medium.

Contamination

The research teams monitored treatment and control classrooms to ensure that treatment group teachers were not sharing curriculum information or materials with teachers in the control group. At research sites with classroom-level random assignment to the treatment and control groups (treatment and control classrooms in the same school or center), the teams' classroom observations indicated that there was little or

no evidence of contamination. There was minimal risk of contamination at sites where pre-kindergarten programs (child care, Head Start centers, or all pre-kindergarten classrooms in an elementary school) were randomly assigned to the treatment or control condition.

Response Rates and Attrition

The baseline data were collected in fall 2003 from the original sample, with an average response rate of 98 percent for the child assessments, 97 percent for the teacher reports, and 84 percent for the parent interviews. For the first follow-up data collection in spring 2004, attrition reduced the percentage of children for whom data were collected to 93 percent of students completing the child assessments, 90 percent having a teacher report, and 79 percent having a parent interview. Further attrition led to an additional decline in the second follow-up data collection in spring 2005, with 85 percent of the original sample completing the child assessments, 72 percent having a teacher report, and 75 percent having a parent interview. Overall, 15 percent of all the students sampled (426 students) were not included in the analyses: 2 percent non-responders during baseline data collection and 13 percent through later attrition. For the individual research teams, the percentage of students sampled who were not included in the analysis ranged from 3 to 34 percent. There was no evidence of differential sample attrition across the treatment and control groups at each research site.

Analysis

Each curriculum was analyzed separately due to the independence of the research teams, the nonrandom assignment of curricula to research teams and sites, and the differences in control conditions. Because students were nested in classrooms or programs and repeatedly assessed with multiple measures, multi-level models containing a series of student, teacher, and classroom-level covariates were used to address the cross-level correlated errors, allowing for a mixture of random and fixed effects (see appendix B for details). For each curriculum, these models were used to estimate differences between treatment and control group means for each of the 27 outcome measures. The type of model used to analyze each outcome measure depended on the number of time points it was observed.

Two types of models for repeated measures (spline and simple) were used for outcome measures with comparable data from two or three time points. Analysis of covariance (ANCOVA) was conducted for outcome measures observed at one time point. The more observations of a measure from different time points included in a model, the better able the model is to identify the parameters of interest, in this case the treatment and control group means of the measures. For this reason, the spline repeated measures model is the preferred model followed by the simple repeated measures model, and then the ANCOVA. The analysis of each measure uses the most preferred model that can be used given the number of time points the measure was observed. Table D lists the model used with each measure.

For the eight student-level outcome measures with observations at three time points, a repeated measures spline model was used to compare the treatment and control group means for the spring pre-kindergarten and spring kindergarten observations. In addition, the model was used to check for differences in group mean measures at the baseline observation, check for such differences at the start of treatment if there was a lag between curriculum implementation and the baseline data collection, and compare the mean rates of growth for the treatment and control groups in pre-kindergarten and in kindergarten (the statistical techniques used are discussed in appendix B and the results from these three analyses are provided in appendix A). For the four student-level outcome measures and five classroom-level outcome measures with observations at two time points, a simple repeated measures model was used to compare the treatment and control group means at spring pre-kindergarten. Similarly, it was used to check on group mean differences at the baseline and start of treatment, and compare the rates of growth in pre-kindergarten.

Table D. Model used with each measure

Outcome	Measure	Times observed	Model
Reading	TERA	3	Spline repeated measures
	WJ Letter Word Identification	3	Spline repeated measures
	WJ Spelling	3	Spline repeated measures
Pre-kindergarten phonological awareness ¹	Pre-CTOPPP	2	Repeated measures
Kindergarten phonological awareness ¹	CTOPP	1	ANCOVA w/ Pre-K baseline
Language	PPVT	3	Spline repeated measures
	TOLD	3	Spline repeated measures
Mathematics	WJ Applied Problems	3	Spline repeated measures
	CMA-A Mathematics Composite	3	Spline repeated measures
	Shape Composition ²	3	Spline repeated measures
Pre-kindergarten behavior ¹	SSRS Social Skills	2	Repeated measures
	SSRS Problem Behavior	2	Repeated measures
	PLBS	2	Repeated measures
Kindergarten behavior ¹	SSRS Social Skills	1	ANCOVA w/ Pre-K baseline
	SSRS Problem Behavior	1	ANCOVA w/ Pre-K baseline
	LBS	1	ANCOVA w/ Pre-K baseline
Classroom quality	ECERS-R	2	Repeated measures
Teacher-child interaction	Arnett Detachment	2	Repeated measures
	Arnett Harshness	2	Repeated measures
	Arnett Permissiveness	2	Repeated measures
	Arnett Positive Interaction	2	Repeated measures
Literacy instruction	TBRS Written Expression	1	ANCOVA
	TBRS Print and Letter Knowledge	1	ANCOVA
Phonological instruction	TBRS Phonological Awareness	1	ANCOVA
Language instruction	TBRS Book Reading	1	ANCOVA
	TBRS Oral Language	1	ANCOVA
Mathematics instruction	TBRS Math Concepts	1	ANCOVA

¹Pre-kindergarten and kindergarten measures are not on the same scale.

²Building Blocks, Shape Composition task

NOTE: ANCOVA: Analysis of covariance. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten). Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

ANCOVA models were used to estimate the difference in mean outcome measures between the treatment and control group in the spring of pre-kindergarten or kindergarten when only one observation was available. The availability of only one observation of a measure occurred in two situations. First, four of the kindergarten student measures (the CTOPP, SSRS Social Skills, SSRS Problem Behaviors, and LBS) were not on the same scales as the pre-kindergarten measures. The ANCOVA model for these kindergarten measures included students' scores on the respective pre-kindergarten scale as a covariate to address any differences in the groups that occurred, despite randomization. Second, six pre-kindergarten classroom instruction measures were based on the TBRS that was given only in the spring of pre-kindergarten. Group mean differences for these were estimated using an ANCOVA without a similar baseline covariate. These models may be biased by any initial differences in instruction that may have existed despite randomization, as there is no baseline measure.

Results

The goal of the PCER initiative was to identify the impact of the 14 preschool curricula on five student-level outcomes (reading, phonological awareness, language, mathematics, and behavior) and six classroom-level outcomes (classroom quality, teacher-child interaction, and four types of instruction). Each outcome was based on one or more of the measures (see table D); thus, the process of determining a curriculum's impact on the outcomes required two steps. First, the models were estimated to identify average differences in the 27 measures between the students receiving the treatment curriculum and those receiving the control and determine whether they were statistically significant. Second, criteria were applied to the set of measures that made up each outcome to determine whether the results for that group of measures showed a finding that the curriculum had an impact on that outcome. This process is described in the following order: (1) the model results for the 27 measures, (2) the criteria applied to the measures for each outcome, and (3) the findings derived from applying the criteria to the results for the measures.

The analysis tested the statistical significance of the difference between the means of the treatment versus the control group for each measure. Tables E-G display this difference as an effect size and note which differences are statistically significant (using a significance level of .05 and a two-tailed test). In the tables, the measures are grouped under their corresponding student-level and classroom-level outcomes. Table E identifies the impacts of each curriculum on the student-level measures in pre-kindergarten (note that *Creative Curriculum* is listed twice as it was implemented by the Vanderbilt University (Tennessee) research team and by the University of North Carolina at Charlotte (North Carolina) research team). Ten curricula show no statistically significant impacts on any of the student-level measures while five show significant impacts on some measures. Table F identifies nine curricula showing no statistically significant impacts on any of the student-level measures in kindergarten and six that do. Table G shows that with seven curricula there are no statistically significant impacts on any of the classroom-level measures and eight curricula show such impacts.

Table E. Effect sizes for student-level measures: Pre-kindergarten

Outcome/Measures	Curricula														
	BB	CC (V)	CC (UNC)	CC with Ldrs	Curiosity Corner	DD	LB	ELLM	LFC	DLM with OC	LE	Pre-K Math	PA	PC	RSL
Reading															
TERA	.39*	.02	-.08	-.30	.10	.06	.02	.15	.16	.68***	.17	.13	.14	.00	.08
WJ Letter Word Identification	.35	.16	-.08	-.16	.09	.10	.10	-.05	.11	.51**	.30	-.01	.42	-.05	.01
WJ Spelling	.18	.19	-.18	.30	.04	.06	.17	.11	.25	.46**	.05	.20	.27	-.15	.20
Phonological awareness															
Pre-CTOPPP	-.07	.10	.02	-.16	.18	.18	-.13	.18	.20	.32*	.14	.04	.05	.10	-.09
Language															
PPVT	.13	.23	.08	-.38	-.01	.15	-.03	.17	.02	.40*	.17	.17	.16	.03	.15
TOLD	.09	.07	-.16	-.22	-.08	.17	.08	.15	.01	.40**	-.04	.17	.15	-.05	-.11
Mathematics															
WJ Applied Problems	.16	.17	.20	-.14	.10	.01	-.10	.10	.20	.36**	.05	.22	.07	.06	.04
CMA-A Mathematics Composite	.14	.10	-.10	.18	.01	.13	.15	.01	.08	.17	-.02	.44**	.18	-.11	-.24*
Shape Composite	-.03	.12	.19	.02	.16	-.13	.21	-.14	.08	.24	-.01	.96***	.27	-.42**	.08
Behavior															
SSRS Social Skills	-.27	.03	.05	-.25	-.06	-.18	-.27	-.06	-.42	-.11	-.06	.22	.04	.22	-.05
SSRS Problem Behavior	.23	.07	-.16	-.01	.43	-.14	-.06	-.24	.37	.11	-.31	-.09	.50	-.08	-.03
PLBS	.04	.14	.07	-.08	-.25	-.18	-.44	.14	-.27	-.16	.17	.09	-.31	.00	.07

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

NOTE: Refer to the glossary for abbreviations of the measures. Abbreviations for the curricula are:

BB: *Bright Beginnings*

CC (V): *Creative Curriculum* (Vanderbilt University)

CC (UNC): *Creative Curriculum* (University of North Carolina at Charlotte)

CC with Ldrs: *Creative Curriculum with Ladders to Literacy*

DD: *Doors to Discovery*

LB: *Let's Begin with the Letter People*

ELLM: *Early Literacy and Learning Model*

LFC: *Language-Focused Curriculum*

DLM with OC: *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

LE: *Literacy Express*

Pre-K Math: *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

PA: *Project Approach*

PC: *Project Construct*

RSL: *Ready, Set, Leap!*

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table F. Effect sizes for student-level measures: Kindergarten

Outcome/Measures	Curricula														
	BB	CC (V)	CC (UNC)	CC with Ldrs	Curiosity Corner	DD	LB	ELLM	LFC	DLM with OC	LE	Pre-K Math	PA	PC	RSL
Reading															
TERA	-.07	.10	-.04	-.54	.43*	-.05	-.13	.30	.05	.76**	-.11	.31	.29	-.03	.01
WJ Letter Word Identification	.09	.38	.00	-.27	.43*	-.09	-.18	.00	.02	.50**	.08	.22	.03	.16	-.12
WJ Spelling	.06	.25	-.05	-.08	.20	-.12	-.06	.04	.11	.22	.06	.03	.14	.00	.04
Phonological awareness															
CTOPP	.01	.06	.06	-.10	.25	-.09	-.13	.08	.03	.38*	.08	-.11	-.17	-.12	-.02
Language															
PPVT	.07	.12	.15	-.30	.14	.18	.00	.34*	-.09	.48**	.16	.11	.10	.10	-.02
TOLD	.16	.11	-.17	-.06	.15	.06	-.12	.44**	-.07	.46**	.10	.08	.32	.01	-.03
Mathematics															
WJ Applied Problems	.13	.17	.09	-.33	.26	-.02	-.13	.26	.11	.48***	-.02	.13	.27	.08	.00
CMA-A Mathematics	.07	.05	.14	-.19	-.05	-.16	-.07	-.05	.00	.13	-.21	.13	.22	-.06	-.10
Composite															
Shape Composite	.15	.00	-.01	-.10	.32	-.12	-.06	.03	.06	.09	-.14	.41***	.24	.12	.03
Behavior															
SSRS Social Skills	.03	.35	-.12	.17	.32	-.05	.24	.27	-.07	-.18	-.37	.06	-.44*	.12	-.03
SSRS Problem Behavior	.24	-.05	.08	.02	-.08	.46	.06	.23	-.05	.01	.22	-.01	.49*	.07	.07
LBS	.30	.08	-.20	-.11	.11	-.32	-.10	.04	.10	-.13	-.38*	.01	-.42*	-.02	-.01

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

NOTE: Refer to the glossary for abbreviations of the measures. Abbreviations for the curricula are:

BB: *Bright Beginnings*

CC (V): *Creative Curriculum* (Vanderbilt University)

CC (UNC): *Creative Curriculum* (University of North Carolina at Charlotte)

CC with Ldrs: *Creative Curriculum with Ladders to Literacy*

DD: *Doors to Discovery*

LB: *Let's Begin with the Letter People*

ELLM: *Early Literacy and Learning Model*

LFC: *Language-Focused Curriculum*

DLM with OC: *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

LE: *Literacy Express*

Pre-K Math: *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

PA: *Project Approach*

PC: *Project Construct*

RSL: *Ready, Set, Leap!*

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table G. Effect sizes for classroom-level measures: Pre-kindergarten

Outcome/Measure	Curricula														
	BB	CC (V)	CC (UNC)	CC with Ldrs	Curiosity Corner	DD	LB	ELLM	LFC	DLM with OC	LE	Pre-K Math	PA	PC	RSL
Global classroom quality ECERS-R	.80	.45	1.66*	-.71	-.48	.39	.82*	-.48	—	.34	1.29*	.05	-.19	.54	.16
Teacher-child interaction															
Arnett Detachment	.19	-.16	-1.68*	.51	-.41	-.07	-.07	-.41	—	-.06	-1.09	-.37	.57	.12	.19
Arnett Harshness	.12	-.12	-.70	-.26	.14	-.38	-.95*	-.40	—	-.70	-.84	.18	.86	-.13	.30
Arnett Permissiveness	.16	.51	-1.01	1.02	-.98	.13	-.05	-.24	—	.05	.51	-.45	-.43	-.02	-.24
Arnett Positive Interactions	.41	-.15	1.65**	.03	.02	.38	.48	.29	—	.43	.56	.16	-.99	.46	.04
Language instruction															
TBRS Book Reading	1.03	-.47	.28	-.32	2.06**	1.18*	.63	.32	-.79	.01	.49	.07	-.76	.81	-.18
TBRS Oral Language	.39	-.07	1.80**	-.50	.37	.59	.44	.14	.87	-.33	.25	.19	-.42	.52	-.24
Phonological instruction															
TBRS Phonological Awareness	1.53*	1.97	-.10	-.19	.44	.58	.66	.53	.92	1.41*	1.26*	.38	-1.19	.01	.22
Literacy instruction															
TBRS Print and Letter Knowledge	1.51*	1.81	1.02	.75	-.99	.90*	.99*	.41	.33	.91	1.07	.07	.34	.34	-.02
TBRS Written Expression	1.61*	1.99	1.73**	1.13*	-.54	.62	.60	-.22	.99	-.58	-.03	-.12	.62	.43	.10
Mathematics instruction															
TBRS Math Concepts	.98	1.48	.75	.44	-.33	.37	.24	-.92	.20	-.46	-.12	.57	-.64	.53	-.10

— Not available.

* $p < .05$; ** $p < .01$

NOTE: Refer to the glossary for abbreviations of the measures. Abbreviations for the curricula are:

BB: *Bright Beginnings*

CC (V): *Creative Curriculum* (Vanderbilt University)

CC (UNC): *Creative Curriculum* (University of North Carolina at Charlotte)

CC with Ldrs: *Creative Curriculum with Ladders to Literacy*

DD: *Doors to Discovery*

LB: *Let's Begin with the Letter People*

ELLM: *Early Literacy and Learning Model*

LFC: *Language-Focused Curriculum*

DLM with OC: *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

LE: *Literacy Express*

Pre-K Math: *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

PA: *Project Approach*

PC: *Project Construct*

RSL: *Ready, Set, Leap!*

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

The statistical significance of these results depend, in part, upon the evaluations having adequate power to detect significant impacts. The original IES Request for Applications to which the 12 research teams successfully responded required that each team include a minimum of 10 classrooms or preschool programs (half treatment and half control) with a minimum of 150 total students. Minimal Detectable Effects were calculated after data collection using the smaller achieved (not expected) samples for each team on a set of four composite measures (combining the measures for reading, language, mathematics and behavior respectively). The Minimal Detectable Effects ranged from .34 to .69 across the composites and teams.

Four of the five student-level outcomes had two to three outcome measures associated with them (phonological awareness only had one per grade), as did three of the six classroom-level outcomes. The measures within an outcome are conceptually related to one another and sufficiently inter-correlated that an effect on one would not be expected to appear, except by chance, without indications of some effect on the others. To minimize the potential for false positive findings that may arise from multiple comparisons made among related measures, a criterion was applied to the set of measures within each outcome (rather than a post-hoc statistical adjustment). These criteria were used to determine whether a curriculum had a treatment effect on each student-level outcome for pre-kindergarten and for kindergarten. They include:

- The reading, mathematics, and behavior outcomes each contained three measures. The finding that a curriculum has an effect on any of these three outcomes required at least two of the three measures to have had a statistically significant effect with the same sign and no significant effect with the opposite sign.
- The language outcome contained two measures. A finding of an outcome effect required at least one of the two measures to have had a statistically significant effect and no significant effect with the opposite sign.
- The phonological awareness outcome contained one measure. A finding of an outcome effect required this measure (Pre-CTOPPP in preschool and CTOPP in kindergarten) to have had a statistically significant effect.

A similar set of rules was used to determine whether a curriculum had a treatment effect on each pre-kindergarten classroom-level outcome:

- The classroom-quality outcome contained one measure. A finding of an outcome effect required this measure to have had a statistically significant effect.
- The teacher-child relationship outcome contained four measures. A finding of an outcome effect required at least two of the four measures to have had a statistically significant effect in the same direction and no statistically significant effects with the opposite direction. For these measures, direction concerns desirability of the effect; a desirable effect would be a positive sign for the Positive Interaction scale and a negative effect for the other three scales.
- The early literacy instruction outcome and the early language instruction outcome each contained two measures. A finding of an outcome effect required at least one of the two measures to have had a statistically significant effect and no significant effect with the opposite sign.
- The phonological instruction outcome and the mathematics instruction outcome each contained one measure. A finding of an outcome effect required the measure to have had a statistically significant effect.

These criteria were applied to the results for each set of measures within the five student-level outcomes (for preschool and for kindergarten) and the six classroom-level outcomes for kindergarten presented in tables E-G. In this way, each curriculum's impact on each of the 16 outcomes was determined. Below, these findings are presented in two sections: the first organized by outcome and the second by curriculum. Under the Findings by Outcome, those curricula affecting each of the five student-level (for pre-kindergarten and

kindergarten) and six classroom-level outcomes (for pre-kindergarten) are identified. Under the Findings by Curriculum, each curriculum is discussed with regard to its effects on the outcomes.

The findings described in both sections are presented in tables H and I. Table H shows the impacts of each curriculum on the student-level outcomes for both pre-kindergarten (pre-K) and kindergarten (K). A blank cell stands for no effect, a plus sign (+) means a positive effect, a minus sign (-) means a negative effect, and a zero (0) signifies no effect in one grade when there is an effect in the other. Table I shows the impact of each curriculum on the classroom-level outcomes using the same symbols.

Findings by Outcome

Two of the 14 intervention curricula had impacts on the student-level outcomes for the pre-kindergarten year (table H). *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* positively affected reading, phonological awareness, and language. *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* curricula positively affected mathematics.

In the kindergarten year, four of the curricula had impacts on the student-level outcomes though three of these did not have impacts during the pre-kindergarten year (table H). *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* continued to have positive effects on reading, phonological awareness, and language in kindergarten as it did in pre-kindergarten. *Curiosity Corner*, which had no effects in pre-kindergarten, was found to positively affect reading in kindergarten. *Early Literacy and Learning Model (ELLM)*, which had no effects in pre-kindergarten, was found to positively affect language in kindergarten. *Project Approach*, which had no effects in pre-kindergarten, was found to negatively affect behavior in kindergarten.

Eight of the 14 treatment curricula had a positive effect on the pre-kindergarten classroom-level outcomes (table I). *Bright Beginnings* affected early literacy instruction and phonological awareness instruction. *Creative Curriculum* (as implemented by the North Carolina team but not by the Tennessee research team) affected classroom quality, teacher-child interaction, early literacy instruction and early language instruction. *Creative Curriculum with Ladders to Literacy* affected early literacy instruction. *Curiosity Corner* affected early language instruction. *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* affected phonological awareness instruction. *Doors to Discovery* affected early literacy instruction and early language instruction. *Let's Begin with the Letter People* affected classroom quality and early literacy instruction. *Literacy Express* affected classroom quality and phonological awareness instruction.

Findings by Curriculum

Each curriculum is discussed separately and cross-curriculum comparisons are not made. The type of pre-kindergarten program involved in the evaluation and the control curricula are described (though the results should not be used to evaluate any control curricula). Impacts on the outcomes are then presented in the following order: (1) student-level outcomes in pre-kindergarten, (2) student-level outcomes in kindergarten, and (3) classroom-level outcomes in pre-kindergarten.

Bright Beginnings

Bright Beginnings and its control were implemented in state pre-kindergarten classrooms in Tennessee. In the control classrooms, teachers used teacher-developed curricula with a focus on basic school readiness. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on early literacy instruction and phonological awareness instruction.

Table H. Findings by student-level outcomes

Curricula	Reading	Phonological awareness	Language	Mathematics	Behavior
<i>Bright Beginnings</i>					
<i>Creative Curriculum (Vanderbilt)</i>					
<i>Creative Curriculum (UNC-Charlotte)</i>					
<i>Creative Curriculum with Ladders to Literacy</i>					
<i>Curiosity Corner</i>	Pre-K: 0 K: +				
<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>	Pre-K: + K: +	Pre-K: + K: +	Pre-K: + K: +		
<i>Doors to Discovery</i>					
<i>Early Literacy and Learning Model</i>			Pre-K: 0 K: +		
<i>Language-Focused Curriculum</i>					
<i>Let's Begin with the Letter People</i>					
<i>Literacy Express</i>					
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>				Pre-K: + K: 0	
<i>Project Approach</i>					Pre-K: 0 K: -
<i>Project Construct</i>					
<i>Ready, Set, Leap!</i>					

NOTE: Abbreviations of the findings are:

Pre-K: Pre-kindergarten

K: Kindergarten

+: Finding of a positive impact

-: Finding of a negative impact

Blank Cell: Finding of no impact

0: Finding of no impact (when an impact is found for the other grade)

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Creative Curriculum—Vanderbilt University

Creative Curriculum and its control were implemented in state pre-kindergarten classrooms in Tennessee. In the control classrooms, teachers used teacher-developed curricula with a focus on basic school readiness. No impacts regarding pre-kindergarten or kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Table I. Findings by classroom-level outcomes

Curricula	Classroom quality	Teacher-child interaction	Early literacy instruction	Phonological awareness instruction	Early language instruction	Math concepts instruction
<i>Bright Beginnings</i>			+	+		
<i>Creative Curriculum</i> (Vanderbilt)						
<i>Creative Curriculum</i> (UNC-Charlotte)	+	+	+		+	
<i>Creative Curriculum with Ladders to Literacy</i>			+			
<i>Curiosity Corner</i>					+	
<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>				+		
<i>Doors to Discovery</i>			+		+	
<i>Early Literacy and Learning Model</i>						
<i>Language-Focused Curriculum</i>						
<i>Let's Begin with the Letter People</i>	+		+			
<i>Literacy Express</i>	+			+		
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>						
<i>Project Approach</i>						
<i>Project Construct</i>						
<i>Ready, Set, Leap!</i>						

NOTE: Abbreviations of the findings are:

+: Finding of a positive impact

Blank Cell: Finding of no impact

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Creative Curriculum—University of North Carolina at Charlotte

Creative Curriculum and its control were implemented in full-day Head Start programs in North Carolina and Georgia. In the control condition, teachers used teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on overall classroom quality, teacher-child relationships, early literacy instruction, and early language instruction.

Creative Curriculum with Ladders to Literacy

Ladders to Literacy was implemented in full-day and half-day Head Start classrooms in New Hampshire as a supplementary curriculum in conjunction with *Creative Curriculum*. In the control condition, teachers used only *Creative Curriculum*. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on early literacy instruction.

Curiosity Corner

Curiosity Corner and its control were implemented in full-day preschool programs in three different states (Florida, Kansas, and New Jersey). In the control condition, teachers used a variety of preschool curricula including the *Creative Curriculum* and *Animated Literacy* curriculum models, and teacher-developed curricula. No impacts regarding pre-kindergarten student-level outcomes were found. A positive impact on reading was found at the end of kindergarten. A positive impact was found at the classroom level on early language instruction.

DLM Early Childhood Express supplemented with Open Court Reading Pre-K

The evaluation of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* took place in public pre-kindergarten classrooms in Florida. In the control condition, teachers were provided with the *High/Scope* curriculum. A positive impact was found on reading, phonological awareness, and language development in both pre-kindergarten and kindergarten. A positive impact was found at the classroom level on phonological awareness instruction.

Doors to Discovery

Doors to Discovery and its control were implemented in full-day Head Start and public pre-kindergarten (Title I and non-Title I) programs in Texas. In the control condition, teachers used teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on early literacy instruction and early language instruction.

Early Literacy and Learning Model (ELLM)

The *Early Literacy and Learning Model (ELLM)* curriculum was implemented in combination with the existing comprehensive curricula that were in use in the control group classrooms in Florida. Several curricula were used in the control classrooms including *Creative Curriculum*, *Beyond Centers and Circletime*, *High Reach*, and *High/Scope*. No impacts regarding pre-kindergarten student-level outcomes were found. A positive impact on language development was found at the end of kindergarten. No impacts were found on the classroom-level outcomes.

Language-Focused Curriculum

The *Language-Focused* curriculum was implemented in full-day Head Start and public pre-kindergarten classrooms in Virginia. The control teachers reported using High/Scope curriculum materials. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. No impacts were found on the classroom instruction outcomes. Impacts on classroom quality and teacher-child interaction outcomes could not be determined because of unreliable (inflated) data from 8 of the 14 participating classrooms on the relevant measures.

Let's Begin with the Letter People

Let's Begin with the Letter People and its control were implemented in full-day Head Start and public pre-kindergarten (Title I and non-Title I) programs in Texas. In the control condition, teachers used teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on classroom quality and early literacy instruction.

Literacy Express

Literacy Express and its control were implemented in public pre-kindergarten classrooms in Florida. In the control condition, teachers were provided with the *High/Scope* curriculum. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on classroom quality and phonological awareness instruction.

Pre-K Mathematics supplemented with DLM Early Childhood Express Math Software

The evaluation of *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* took place in Head Start and public pre-kindergarten classrooms in California and New York. Several curricula were used

in the control condition including *Creative Curriculum*, *High/Scope*, *Montessori*, specialized literacy curricula, and local school district and teacher-developed curricula. A positive impact was found on students' mathematical knowledge at the end of pre-kindergarten. No impacts on the kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Project Approach

The *Project Approach* curriculum was implemented in public pre-kindergarten classrooms in Wisconsin. In the control classrooms, teachers reported implementing their own teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten student-level outcomes were found. A negative impact on behavior was found at the end of kindergarten. No impacts were found on the classroom-level outcomes.

Project Construct

Project Construct was implemented in full-day child care centers in Missouri. In the control schools, teacher-developed generic curricula were implemented. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Ready, Set, Leap!

Ready, Set, Leap! was implemented in pre-kindergarten programs in New Jersey. In the control condition, teachers used the *High/Scope* approach. No impacts on the pre-kindergarten and kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Chapter 1. An Overview of the Preschool Curriculum Evaluation Research Initiative

In 2002, the Institute of Education Sciences (IES) began the Preschool Curriculum Evaluation Research (PCER) initiative to conduct rigorous efficacy evaluations of available preschool curricula. Twelve research teams implemented one or two curricula in preschool settings serving predominantly low-income children under an experimental design. For each team, preschools or classrooms were randomly assigned to the intervention curricula or control curricula and the children were followed from pre-kindergarten through kindergarten. RTI International (RTI) and Mathematica Policy Research (MPR) evaluated the impact of each of the 14 curricula implemented using a common set of measures with the cohort of children beginning preschool in the summer-fall of 2003. This chapter describes the background to the PCER initiative and details the common elements of the evaluations including the experimental design, implementation, analysis, results, and findings.

Study Background

Despite decades of federal, state, and local programs intended to support young children's preparation for schooling, children from low-income families continue to begin formal schooling at a disadvantage. These differences in reading and mathematics achievement based on poverty status are evident at the beginning of kindergarten and persist throughout the elementary years (National Research Council 2001). For example, findings from the Early Childhood Longitudinal Study (ECLS), a multiyear study following a nationally representative sample of more than 22,000 children in the kindergarten class of 1998 through the primary grades, show that children from families living in poverty continue to have lower reading and mathematics achievement scores, on average, than students living in households at or above the poverty line (Princiotta, Flanagan, and Germino-Hausken 2006; West, Denton, and Reaney 2001). At the time of the ECLS fifth-grade follow-up in the spring of 2004, 61 percent of students in poverty scored in the lowest third of the distribution of reading achievement scores, compared to 25 percent of students in households at or above the poverty threshold. In mathematics, 57 percent of students in poverty scored in the lowest third of the distribution of mathematics achievement scores, compared to 26 percent of students in households at or above the poverty threshold. In short, substantial numbers of children from low-income families begin kindergarten behind their more affluent peers, and remain behind as they continue through school.

School Readiness and Later Academic Achievement

Children's early performance in both academic and social domains has been associated with later academic and social outcomes as they make the transition from preschool to formal instruction in kindergarten and first grade (Downer and Pianta 2006; Miles and Stipek 2006). Research has found stability in children's early language and literacy skills and abilities (Dickinson and Tabors 2001; Entwisle and Alexander 1988; Hart and Risley 1995). Children who enter kindergarten with poor language and literacy skills tend to show poor reading achievement during the early grades, and this relatively poor reading performance tends to be maintained into early and late adolescence (Cunningham and Stanovich 1997; Cunningham, Stanovich, and West 1994; Echols et al. 1996; Juel 1988; Lentz 1988; Stanovich 1986). In contrast, children who begin formal schooling with strong emergent literacy skills learn to read earlier and develop better reading skills, thus providing a foundation for later academic competence (Downer and Pianta 2006; Princiotta, Flanagan, and Germino-Hausken 2006). Phonological awareness has also been related to general reading ability (Chaney 1992; Ehri and Wilce 1980; Liberman et al. 1974; Perfetti et al. 1987; Shankweiler et al. 1995) and there is an

association between children's phonological awareness skills in kindergarten and their reading achievement in later years of school (Juel 1991; Scarborough 1989; Stanovich 1986; Wagner, Torgeson, and Rashotte 1994).

Early understanding of mathematics concepts during preschool is similarly important. Recent research has revealed a relationship between the extent of young children's mathematical knowledge and mathematics achievement in school (Duncan et al. 2006; Entwisle and Alexander 1992; Natriello, McDill, and Pallas 1990). Children from low-income families perform below their middle-income peers on national and international mathematics assessments as early as the preschool years and these gaps in performance can persist into the elementary school grades (Duncan et al. 2006; Entwisle and Alexander 1992) and into early and late adolescence (Downer and Pianta 2006; Perie, Grigg, and Donahue 2005). For example, in the 2005 National Assessment of Educational Progress, 33 percent of fourth-grade children from low-income families performed below the basic level as compared to 10 percent of children from other socioeconomic backgrounds (Perie, Grigg, and Donahue 2005).

In addition to early language, literacy, and mathematical knowledge, children's behavior (including early social skills) has also been associated with both early and later school success (Downer and Pianta 2006; Miles and Stipek 2006). For example, prosocial behavior and social competence predict academic performance in the early grades, whereas childhood aggression is increasingly associated with school failure later in elementary school (Miles and Stipek 2006). As Zins et al. (2004) note, learning is a social process, and problems following directions, or difficulties getting along with others and controlling negative emotions, distract from learning.

Early Childhood Education

A potential avenue for improving school readiness among young children at risk for school failure is through early childhood education. As recently as 2005, almost half (47%) of all children aged 3-5 years from low-income families were enrolled in either part-day or full-time early childhood programs (U.S. Department of Education 2006). A variety of preschool curricula are in use in these early childhood programs. There is little information based on rigorous evaluation regarding which of these curricula are most effective for improving children's school readiness as defined by pre-reading skills, language skills, early mathematics knowledge, and behavioral skills.

In the past, rigorous evaluation research has focused on model demonstration programs such as the Abecedarian and Perry Preschool programs. In the Abecedarian program children enrolled as infants received intervention services for 6 to 8 hours a day, 5 days per week, 50 weeks per year. Intervention continued through the first 3 years of school, with a resource teacher working with each child and family to support their regular schooling (Campbell et al. 2002). The Abecedarian project was an intensive, long-term early intervention for young children that does not reflect typical practice in the early childhood programs in place today.

The Perry Preschool project included half-day sessions 5 days a week for 2 academic years along with weekly home visits by the teachers to involve mothers and their children in educational activities in the home. The preschool program component of the Perry Preschool project is more similar to current early childhood interventions than the Abecedarian project. However, it differs from an evaluation of contemporary preschool in two ways. First, the participants represented a restricted population of preschool children—the children were African American children from low-income families who had low IQ scores (70-85, which is the range for the educable mentally retarded) (Schweinhart 2004). Second, the study was designed to compare a treated group (i.e., children who received early childhood center-based program services and home visits) to an untreated group (no early childhood center-based program or home visits). A further limitation to the generalizability of the study is that the sample was small—only 123 children.

Although both the Abecedarian and Perry Preschool programs have shown long-term benefits for participants, it is not clear that findings from their evaluations are directly applicable to less intensive, school-based early childhood programs for typically developing children that are in place today.

The Preschool Curriculum Evaluation Research Initiative

The lack of relevant evidence of the impact of current preschool curricula on children's school readiness led IES to begin the PCER initiative in 2002. Rigorous efficacy evaluations were to be conducted on preschool curricula using a randomized experimental design to ensure that any systematic difference found between the treatment and control groups was due to the intervention curricula. The student-level outcomes of greatest interest were those skills that are highly predictive of academic success in the early years of elementary school and influenced by curricula and practice.

Under a competitive process, 12 research teams received peer-reviewed grants to implement one to two preschool curricula of their choosing with a predominantly low-income population under an experimental design. Teams were required to include a minimum of 10 classrooms or preschool programs (half treatment and half control) and 150 students. Under the Request for Applications, teams were asked to propose preschool curricula with sufficient standardized training procedures and published materials to support implementation of the curriculum by entities other than the curriculum developer. The set of curricula evaluated was determined by the grants awarded.

Contracts were awarded to RTI and MPR to individually evaluate the 14 preschool curricula using a common battery of measures. One cohort of students was to be followed from the start of preschool in the fall of 2003 through the end of kindergarten in the spring of 2005. Data collection included child assessments, parent interviews, teacher reports on children's social skills, teacher interviews and questionnaires, and direct classroom observations (preschool year only).

Research Questions

The PCER initiative focused on the impact of the intervention curricula on students' reading, phonological awareness, early language, early mathematics knowledge, and behavior (including social skills) at the end of pre-kindergarten and kindergarten. As described above, these domains of knowledge and skills are predictive of academic success in the early years of elementary school.

In addition, the PCER evaluation study also examined the impact of the curriculum interventions on teachers' classroom instructional practice, teacher-child interaction, and global classroom quality. These dimensions of early childhood programs have been posited as mediators (e.g., instructional practice) and moderators (e.g., teacher-child interaction, classroom quality) of the relation between early childhood curricula and child outcomes (Arnett 1989; Peisner-Feinberg and Burchinal 1997; Ruopp et al. 1979).

In sum, the research questions for the evaluation primarily concern student academic and behavioral outcomes and also include classroom outcomes due to their potentially mediating or moderating roles. The research questions are:

1. What is the impact of each of the 14 preschool curricula on preschool students' reading skills, phonological awareness, language development, mathematical knowledge, and behavior?
2. What is the impact of each of the 14 preschool curricula on these outcomes for students at the end of kindergarten?
3. What is the impact of each of the 14 preschool curricula on preschool classroom quality, teacher-child interactions, and instructional practices?

Study Design

The PCER evaluation study is composed of a set of individual evaluations of 14 pre-kindergarten curricula in which each of the 12 research teams selected and implemented curricula at each of their research sites. All research teams identified pre-kindergarten programs serving children from low-income families and recruited the programs, teachers, parents, and children for participation in a random assignment study to evaluate the chosen curriculum or curricula. Within the evaluation for each team, participating schools or classrooms were randomly assigned to treatment or control group conditions. Each research team provided training and support to the treatment group teachers who implemented the curriculum at their research site. Control group teachers were instructed to continue using the prevailing curriculum that was in use in their classroom prior to the start of the evaluation study. RTI and MPR evaluated the impact of each curriculum using a common set of measures.

Rather than one overall evaluation, the PCER study contains individual evaluations for each curriculum for three reasons. First, each research team worked independently. Second, the selection of the intervention and the randomized assignment occurred at the team level. Third, different control curricula were used with each intervention curriculum. The findings from the evaluations will determine whether a curriculum was more effective at its research site than the control curriculum used there. The findings cannot determine the effectiveness of the intervention curricula in relation to one another.

Intervention and Control Curricula

The 12 research teams were responsible for selecting the curricula that they implemented and would be evaluated by either RTI or MPR. The curricula, corresponding research team, research site, and evaluator are listed in table 1.1. Three teams each implemented two curricula. Two teams implemented the same curriculum, *Creative Curriculum*. Four teams had originally developed the curricula that they implemented (*Curiosity Corner*, *Literacy Express*, *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*, and *Early Literacy and Learning Model*). RTI evaluated eight curricula implemented by seven teams (including one curriculum which was evaluated by two teams) while MPR evaluated six curricula implemented by five teams. In sum, 14 different curricula (one twice) were evaluated.

The 14 curricula were evaluated in comparison with the local control condition that, in general, was the local curriculum-as-usual. As a result, multiple curricula were used across the control sites and within some of the individual evaluations. These included teacher-developed nonspecific curricula with a focus on basic school readiness, district-developed curricula, and published curricula (some of which were implemented by other research teams). Table 1.2 matches the intervention curricula with their control curricula. The control curricula fully differed from the intervention curricula except in two cases in which the intervention curriculum was an add-on to the existing curriculum. For the University of New Hampshire (New Hampshire) research team, *Ladders to Literacy* was implemented as a supplementary curriculum to *Creative Curriculum* and the latter was the control condition. For the research team from the University of California, Berkeley and the University at Buffalo, State University of New York (California/New York research team), the *Pre-K Mathematics with DLM Early Childhood Express Math software* was added on to the existing curricula in use. In addition, *Creative Curriculum* was implemented by two teams but was also the control for two other teams. Because different control curricula were used among the evaluations, this report does not make cross-intervention comparisons.

Table 1.1. The intervention curricula

Curriculum and publisher	Research team	Research site	Evaluator
<i>Bright Beginnings</i> (Charlotte-Mecklenburg Schools 2001)	Vanderbilt University	Tennessee	RTI
<i>Creative Curriculum</i> (Teaching Strategies, Inc. 2002)	Vanderbilt University	Tennessee	RTI
<i>Creative Curriculum</i> (Teaching Strategies, Inc. 2002)	University of North Carolina at Charlotte	North Carolina and Georgia	RTI
<i>Creative Curriculum with Ladders to Literacy</i> (Teaching Strategies, Inc. 2002; Paul H. Brookes Publishing Company 1998)	University of New Hampshire	New Hampshire	RTI
<i>Curiosity Corner</i> (Success for All Foundation, Inc. 2003)	Success for All Foundation	Florida, Kansas, New Jersey	MPR
<i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i> (SRA/McGraw-Hill 2003)	Florida State University	Florida	MPR
<i>Doors to Discovery</i> (Wright Group/McGraw-Hill 2001)	University of Texas Health Science Center at Houston	Texas	RTI
<i>Early Literacy and Learning Model</i> (Florida Institute of Education and the University of North Florida 2002)	University of North Florida	Florida	RTI
<i>Language-Focused Curriculum</i> (Paul H. Brookes Publishing Company 1995)	University of Virginia	Virginia	MPR
<i>Let's Begin with the Letter People</i> (Abrams & Company 2000)	University of Texas Health Science Center at Houston	Texas	RTI
<i>Literacy Express</i> (unpublished) (Author: Lonigan and Farver 2002, unpublished)	Florida State University	Florida	MPR
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i> (Scott Foresman - Pre-K Mathematics 2002; SRA/ McGraw-Hill - DLM Early Childhood Express Math software 2003)	University of California, Berkeley and University at Buffalo, State University of New York	California and New York	RTI
<i>Project Approach</i> (Ablex 1989)	Purdue University and University of WI-Milwaukee	Wisconsin	RTI
<i>Project Construct</i> (Missouri Department of Elementary and Secondary Education 1992)	University of Missouri- Columbia	Missouri	MPR
<i>Ready, Set, Leap!</i> (LeapFrog School House 2003)	University of California, Berkeley	New Jersey	MPR

NOTE: RTI: RTI International

MPR: Mathematica Policy Research, Inc.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table 1.2. The intervention and control curricula

Intervention curriculum	Research site	Control curriculum
<i>Bright Beginnings</i>	Tennessee	"Homegrown" nonspecific curricula
<i>Creative Curriculum</i> (Vanderbilt)	Tennessee	"Homegrown" nonspecific curricula
<i>Creative Curriculum</i> (UNC-Charlotte)	North Carolina and Georgia	"Homegrown" nonspecific curricula
<i>Creative Curriculum with Ladders to Literacy</i>	New Hampshire	<i>Creative Curriculum</i>
<i>Curiosity Corner</i>	Florida, Kansas, New Jersey	<i>Creative Curriculum and Animated Literacy</i>
<i>DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i>	Florida	<i>High/Scope Curriculum</i>
<i>Doors to Discovery</i>	Texas	"Homegrown" nonspecific curricula
<i>Early Literacy and Learning Model</i>	Florida	<i>Creative Curriculum, Beyond Centers and Circletime, High Reach, or High/Scope</i>
<i>Language-Focused Curriculum</i>	Virginia	<i>High/Scope Curriculum</i>
<i>Let's Begin with the Letter People</i>	Texas	"Homegrown" nonspecific curricula
<i>Literacy Express</i>	Florida	<i>High/Scope Curriculum</i>
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i>	California and New York	CA: Various "homegrown" and <i>High/Scope</i> NY: <i>Creative Curriculum</i> and Buffalo Public Schools Benchmarks
<i>Project Approach</i>	Wisconsin	"Homegrown" nonspecific curricula
<i>Project Construct</i>	Missouri	Teacher-developed generic curriculum
<i>Ready, Set, Leap!</i>	New Jersey	" <i>High/Scope</i> philosophy"

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Sample and Random Assignment to Condition

Preschool programs taking part in the evaluation of the curricula included Head Start centers, private childcare centers, and public pre-kindergarten programs in urban, rural, and suburban locations. Each research team recruited interested local preschool programs. As required by IES, the research teams selected preschool programs serving children from low-income families. Programs agreed to the random assignment (by program or classroom) to a treatment curriculum or to local control conditions.

For each evaluated curriculum, table 1.3 indicates whether pre-kindergarten programs or classrooms were randomly assigned to treatment or control conditions, the number assigned to each, and the number of treatment and control students included in each evaluation. Three teams (implementing four curricula) used randomly assigned pre-kindergarten programs and the other nine teams used randomly assigned classrooms. Three teams compared two curricula against a single set of control classrooms or programs. Across all the teams, 2,911 children, 315 preschool classrooms, and 208 preschools from a total of 16 different geographical locations were part of the curricula evaluations.

Table 1.3. Units of random assignment for evaluation of each curriculum

Research team	Curricula	Treatment sample	Control sample	Students
Vanderbilt University	<i>Bright Beginnings</i>	7 classrooms	7 classrooms	T: 103
	<i>Creative Curriculum</i>	7 classrooms		C: 105
University of North Carolina at Charlotte	<i>Creative Curriculum</i>	9 classrooms	9 classrooms	T: 97 C: 97
University of New Hampshire	<i>Creative Curriculum with Ladders to Literacy</i>	7 classrooms	7 classrooms	T: 62 C: 61
Success for All Foundation	<i>Curiosity Corner</i>	10 Pre-K programs	8 Pre-K programs	T: 105 C: 110
University of Texas Health Science Center at Houston	<i>Doors to Discovery</i>	14 classrooms	15 classrooms	T: 101
	<i>Let's Begin with the Letter People</i>	15 classrooms		C: 96
University of North Florida	<i>Early Literacy and Learning Model</i>	14 classrooms ¹	14 classrooms ¹	T: 137 C: 107
University of Virginia	<i>Language-Focused Curriculum</i>	7 classrooms	7 classrooms	T: 97 C: 98
Florida State University	<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>	5 Pre-K programs	6 Pre-K programs	T: 101
	<i>Literacy Express</i>	6 Pre-K programs		C: 97
UC-Berkeley and University at Buffalo, State University of New York	<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>	20 classrooms	20 classrooms	T: 159 C: 157
Purdue University and University of WI-Milwaukee	<i>Project Approach</i>	7 classrooms	6 classrooms	T: 114 C: 90
University of Missouri-Columbia	<i>Project Construct</i>	10 Pre-K programs ¹	11 Pre-K programs ¹	T: 123 C: 108
UC-Berkeley	<i>Ready, Set, Leap!</i>	18 classrooms	21 classrooms	T: 149 C: 137

¹ After one program or classroom attrited.

NOTE: T: Treatment Group

C: Control Group

Three research teams (Vanderbilt University, University of Texas Health Science Center at Houston, and Florida State University) have two treatment groups and a shared control group. When reading the "Students" column, the first "T" refers to the first curriculum in the same row, while the second "T" refers to the second curriculum in the same row. The "C" refers to the shared control group. For example, Vanderbilt University compared two curricula: *Bright Beginnings* (103 students) and *Creative Curriculum* (101 students) to a control curriculum (105 students).

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

The process of random assignment differed somewhat depending upon the evaluator. The seven research teams working with RTI were responsible for the random assignment at their sites and RTI monitored the process and tracked any changes. Teams monitored the assignment of children to classrooms and reported that there was no evidence of preferential assignment of children to treatment and control group status. These teams had a pilot preschool implementation year starting in the fall of 2002. All teams randomized classrooms and all but two (the Purdue University and University of Wisconsin-Milwaukee [Purdue/Wisconsin] research team and the New Hampshire research team) used block random assignment. Blocking differed by team and included demographics (e.g., similar neighborhoods or schools), type of preschool program (e.g., Head Start or public preschool), feeder elementary school performance, and teacher qualifications (e.g., education level and certification). The randomization done in the pilot year carried over to the actual evaluation begun in the 2003-04 school year, with some modifications. Along with a new student cohort, the evaluation year also saw changes in teachers and classrooms from the pilot year. Teacher turnover occurred for all teams and was purposely high for two of them. The Purdue/Wisconsin research team recruited all new treatment teachers to avoid a mix of first- and second-year implementers (and re-randomized all teachers) while the other teams retained a majority of their treatment teachers. There were some changes in classrooms requiring new classrooms to be randomized into the treatment and control groups. The University of North Florida (Florida-UNF) research team randomly selected all new control classrooms because the pilot year control teachers were trained in the treatment curriculum for another study. The Tennessee research team replaced eight classrooms, the University of California, Berkeley with the University at Buffalo, State University of New York (California/New York) research team replaced three classrooms, the University of North Florida research team (Florida-UNF) replaced one classroom, the New Hampshire team added two classrooms, and the University of North Carolina at Charlotte (North Carolina) research team dropped two classrooms because of their participation in a program to improve pre-kindergarten provided by the state of North Carolina.

For the five research teams working with MPR, randomization was done at the beginning of the preschool evaluation year. MPR in conjunction with the teams conducted block random assignment for four of them and Florida State University (FSU) block randomly assigned pre-kindergarten programs to its two curricula and control. Assignment to treatment or control was done by preschool for three teams (implementing four curricula) and by classroom for two teams. To increase the precision in estimating program impacts, classrooms or schools were grouped into blocks of two or more based on such characteristics as teacher's experience, school location, or state performance score. For school-level assignment, MPR sorted the preschools by block and assigned a random number (using a function in MS Excel) to each. Within each block, the highest numbered preschool was assigned to treatment and the next to control, and this process was repeated until all preschools were assigned. For classroom-level assignment, the same procedure was used with classrooms sorted by block. The Florida State University (FSU) research team blocked preschools by a state letter grade (A-D) school rating system and within the ratings, ranked the preschools by teacher experience. Starting at the top of this ranking system (experience within grades), schools were grouped into triplets. The three preschools within each triplet were randomly assigned with one assigned to the first intervention curriculum, one going to the second intervention curriculum, and one going to the control.

Kindergarten sample dispersal

In the follow-up year of the study (2004-05), the preschool sample of children dispersed into a total of 1,513 kindergarten classrooms and 868 schools. The students' exposure to the treatment curriculum and their teachers' training in its use did not carry over to their kindergarten year except in one case. In the design for the evaluation of *Curiosity Corner*, some students from each preschool were to attend kindergartens using the SFA *Kinder Corner* curriculum while others would attend kindergartens not using it. Table 1.4 provides a summary of the transition of each research team's sample from the preschool classrooms and schools into the kindergarten classrooms and schools.

Table 1.4. Dispersion of the preschool study sample into kindergarten schools and classrooms

Research team (Curricula)	Preschool		Kindergarten	
	Number of classrooms	Number of schools	Number of classrooms	Number of schools
Total	315	208	1,513	868
Vanderbilt University (<i>Bright Beginnings; Creative Curriculum</i>)	21	19	134	64
University of North Carolina at Charlotte (<i>Creative Curriculum</i>)	18	5	122	54
University of New Hampshire (<i>Creative Curriculum with Ladders to Literacy</i>)	14	8	41	26
Success for All Foundation (<i>Curiosity Corner</i>)	31	18	107	69
University of Texas Health Science Center at Houston (<i>Doors to Discovery; Let's Begin with the Letter People</i>)	44	19	149	78
University of North Florida (<i>ELLM</i>)	28	28	175	119
University of Virginia (<i>Language-Focused Curriculum</i>)	14	5	54	21
Florida State University (<i>Literacy Express; DLM Early Childhood Express supplemented with Open Court Reading Pre-K</i>)	30	17	145	46
University of California, Berkeley and University at Buffalo, SUNY (<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software</i>)	40	35	200	136
Purdue University and University of Wisconsin-Milwaukee (<i>Project Approach</i>)	13	12	58	37
University of Missouri (<i>Project Construct</i>)	23	21	166	124
University of California, Berkeley (<i>Ready, Set, Leap!</i>)	39	21	162	94

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Measures

A common set of measures was used with each research team. The measures were chosen for two purposes. First, some of the measures provided descriptive data on the students, teachers, and parents to be used as background information; determined whether the groups included were those targeted by the PCER evaluation study; checked whether the randomization process succeeded in providing similar treatment and control groups; and created variables that should be controlled for in the statistical analysis because they are known to be related to student achievement.

Second, 27 measures were chosen to address the outcomes of interest regarding children's school readiness (reading, phonological awareness, language, mathematics and behavior) and classroom conditions (classroom quality, teacher-child interaction, and instructional practices). Table 1.5 lists the measures used for each outcome, when they were collected, and through which instrument they were collected. Five major data collection instruments were used to collect the outcome measures and other student, school, and family data: (1) a child assessment, (2) a teacher report, (3) classroom observation, (4) a teacher interview or questionnaire, and (5) a parent interview. Each instrument and the measures it included are discussed below. Information on the measures is derived from the data available in the published technical manuals and includes reliabilities and age appropriateness of the test. For the measures developed for the PCER study, the information presented has not been published and was provided by the PCER Evaluation Consortium. In addition to the published reliabilities for each measure provided in the text, table 1.6 provides the reliabilities of each measure (based on internal consistency) calculated from the PCER data. Table 1.6 also provides the publisher's standardized score scale for each measure, or the raw score range for measures that do not use a standardized scale.

Child assessment

The child assessment contained 10 measures of four student academic outcomes: reading, phonological awareness, language, and mathematics. Each measure was individually given as a preschool pre-test in the fall of 2003 and as post-tests near the end of preschool in the spring of 2004 and the end of kindergarten in the spring of 2005. One exception was that the Pre-CTOPPP was given as the preschool pre- and post-test and the CTOPP was given as the kindergarten post-test.

The child assessment used a combination of commercially available and in-development measures. The former have been standardized and subject to reliability and validity studies (see references for each measure below). The latter have not but were included because they addressed relevant knowledge and skills for which more developed and validated measures for preschool children were not available. Results for the less developed measures should be interpreted with caution. The average length of the child assessments was 61.5 minutes (with a standard deviation [SD] of 16.6 minutes) in the fall of 2003, 61.9 (SD = 14.7) minutes in spring 2004, and 75.7 (SD = 19.2) minutes in the spring of 2005. Child assessments that were longer than 45 minutes were generally completed in two assessment sessions. The child assessment measures used in the evaluation of the curricula include:

Early reading measures

- a. **Test of Early Reading Ability, 3rd Edition (TERA-3):** The TERA-3 is a standardized measure of children's mastery of early, developing reading skills (Reid, Hresko, and Hammill 2001). It includes three subtests: alphabet, conventions, and meaning. The alphabet subtest measures knowledge of the alphabet and correspondence between sounds and letters, knowledge of letter names, the ability to determine the initial and final sounds in printed words, knowledge of the number of sounds and syllables in printed words, and the awareness of letters printed in different forms. The conventions subtest measures book handling (e.g., knowing the correct orientation of a book, where to begin reading, and where the top and bottom of the page are); print conventions (e.g., letter orientation, case, presentation of print, text genre, and knowledge of punctuation, capitalization, and spelling). The meanings subtest measures the ability to comprehend the meaning of printed material by presenting children with pictures of labeled common objects and simple words, and asking them to point to words or read simple words and phrases. Subtests are standardized to have a mean of 10 and a standard deviation of 3, and the reading composite has a mean of 100 and a standard deviation of 15. The reliability of the reading composite has been evaluated through internal consistency (.91-.97) and test-retest (.98). The test is appropriate for students aged 3 years and 6 months to 8 years and 6 months. Administration can take 15 to 45 minutes.

Table 1.5. Outcomes and measures

Outcome	Measure	Times collected	Instrument
Reading	TERA	Pre-K: fall/spring, K: spring	Child assessment
	WJ Letter Word Identification	Pre-K: fall/spring, K: spring	
	WJ Spelling	Pre-K: fall/spring, K: spring	
Phonological awareness ¹	Pre-CTOPPP CTOPP	Pre-K: fall/spring, K: spring	Child assessment
Language	PPVT	Pre-K: fall/spring, K: spring	Child assessment
	TOLD	Pre-K: fall/spring, K: spring	
Mathematics	WJ Applied Problems	Pre-K: fall/spring, K: spring	Child assessment
	CMA-A	Pre-K: fall/spring, K: spring	
	Shape Composition ²	Pre-K: fall/spring, K: spring	
Pre-kindergarten behavior ¹	SSRS Social Skills	Pre-K: fall/spring	Teacher report
	SSRS Problem Behavior	Pre-K: fall/spring	
	PLBS	Pre-K: fall/spring	
Kindergarten behavior ¹	SSRS Social Skills	K: spring	Teacher report
	SSRS Problem Behavior	K: spring	
	LBS	K: spring	
Classroom quality	ECERS-R	Pre-K: fall/spring	Classroom observation
Teacher-child interaction	Arnett Detachment	Pre-K: fall/spring	Classroom observation
	Arnett Harshness	Pre-K: fall/spring	
	Arnett Permissiveness	Pre-K: fall/spring	
	Arnett Positive Interaction	Pre-K: fall/spring	
Literacy instruction	TBRS Written Expression	Pre-K: spring	Classroom observation
	TBRS Print and Letter Knowledge	Pre-K: spring	
Phonological instruction	TBRS Phonological Awareness	Pre-K: spring	Classroom observation
Language instruction	TBRS Book Reading	Pre-K: spring	Classroom observation
	TBRS Oral Language	Pre-K: spring	
Mathematics instruction	TBRS Math Concepts	Pre-K: spring	Classroom observation

¹Pre-kindergarten and kindergarten measures are not on the same scale.

²Building Blocks, Shape Composition task

NOTE: Pre-K: Pre-kindergarten

K: Kindergarten

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

- b. **Woodcock-Johnson Letter Word Identification:** This is a standardized measure of identification of letters and reading of words (McGrew and Woodcock 2001). It has a mean of 100 and a standard deviation of 15. Its reliability has been evaluated through test-retest (.87-.96). The test is appropriate for students aged 2 years and older. Administration takes 5 minutes.

Table 1.6. Standardized mean and reliability for outcome measures

Measure	Standardized mean (standard deviation)	Internal consistency (Cronbach's Alpha)		
		Fall Pre-K	Spring Pre-K	Kindergarten
TERA	100 (15)	.91	.94	.94
		.90	.93	.88
WJ Letter Word Identification	100 (15)	.86	.89	.92
		.87	.90	.93
WJ Spelling	100 (15)	.86	.88	.85
		.85	.83	.81
Pre-CTOPPP	Not standardized: 0-18 score	.83	.85	†
		.83	.88	†
CTOPP	10 (3)	†	†	.37
		†	†	.88
PPVT	100 (15)	.96	.96	.95
		.96	.96	.95
TOLD	10 (3)	.82	.79	.73
		.86	.85	.80
WJ Applied Problems	100 (15)	.85	.83	.81
		.75	.79	.75
CMA-A	Not standardized: 0-1 composite score	.79	.78	.70
		.76	.75	.75
Shape Composition ¹	Not standardized: 0-1 score	—	—	—
SSRS Social Skills	100 (15)	.95	.94	.95
		.94	.95	.94
SSRS Problem Behaviors	100 (15)	.86	.86	.91
		.84	.85	.91
PLBS	50 (10)	.92	.92	†
		.91	.93	†
LBS	50 (10)	†	†	.91
		†	†	.92
ECERS-R	Not standardized: 1-7 score	.93	.94	†
Arnett Detachment	Not standardized: 1-4 score	.63	.80	†
Arnett Harshness	Not standardized: 1-4 score	.78	.85	†
Arnett Permissiveness	Not standardized: 1-4 score	.50	.62	†
Arnett Positive Interaction	Not standardized: 1-4 score	.86	.86	†
TBRS	Not standardized: 0-7 score	†	†	†

— Not available.

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: Reliabilities calculated by Mathematica Policy Research, Inc. (shaded) and RTI International (not shaded).

No reliabilities calculated for the Shape Composition or the TBRS measures. When a measure was used in only one grade, the cells for the other grades are marked. Refer to the glossary for abbreviations of the measures.

SOURCE: The Child Assessment (Fall 2003, Spring 2004, and Spring 2005); The Teacher Child Report data (Fall 2003, Spring 2004, and Spring 2005); The Classroom Observation data (Fall 2003 and Spring 2004).

- c. **Woodcock-Johnson Spelling:** This is a standardized measure that assesses children's prewriting skills, such as drawing lines and tracing, writing letters, and spelling of orally presented words (McGrew and Woodcock 2001). It has a mean of 100 and a standard deviation of 15. Median reliability is .89 for students aged 5 to 19 years. The test is appropriate for students 2 years and older. Administration takes 5 minutes.

Phonological awareness measures

- a. **Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, Pre-kindergarten:** This subtest assesses children's phonological awareness (i.e., ability to identify and manipulate sounds in spoken words). It uses word props and picture plates for the first nine items to help younger children understand the task. During the evaluations, this measure was still in research version form and standardization had not yet been completed so raw scores were used on a scale of 0-18. A commercially available version, known as the Test of Preschool Early Literacy, was released afterward for students aged 3-5 years (or older) with a testing time of 25 to 30 minutes (Lonigan et al. 2002).
- b. **Comprehensive Test of Phonological Processing (CTOPP), Elision subtest, Kindergarten:** The CTOPP-Elision subtest assesses phonological awareness and is similar to the Pre-CTOPPP-Elision subtest but does not include pictures in the administration format (Wagner, Torgeson, and Rashotte 1999). It has been standardized to have a mean of 10 and standard deviation of 3. Its reliability has been measured through internal consistency (.90-.91) and test-retest after 2 weeks (.88). However, RTI calculated a low reliability of .37 (see table 1.6) for all the research teams it worked with raising cautions when interpreting the impact analysis results for the CTOPP. The CTOPP version administered is appropriate for students aged 5 and 6 years. It was given in the spring of the kindergarten year and administration took 5 to 10 minutes.

Language measures

- a. **Peabody Picture Vocabulary Test, 3rd Edition (PPVT-III):** The PPVT is a standardized measure of children's receptive vocabulary that has also been used to estimate children's cognitive ability (Dunn and Dunn 1997; Williams and Wang 1997). It measures the child's knowledge of the meaning of spoken words and his or her receptive vocabulary for standard American English. The child is not required to define words but to show understanding of what they mean by pointing to a picture that best represents the meaning. The difficulty level of the PPVT test ranges from easy for children aged 2.5 years to difficult for adults. PPVT-III has been standardized to have a mean of 100 and standard deviation of 15. Its reliability has been evaluated through internal consistency (.93-.95), split-half (.90-.94), and test-retest (.91-.94). Administration takes 11 to 12 minutes.
- b. **Grammatical Understanding subtest from Test of Language Development-Primary: 3rd Edition (TOLD-P:3), Grammatical Understanding subtest:** The Grammatical Understanding subtest as used in the PCER evaluation has 25 items to assess the child's ability to comprehend the meaning of sentences (Newcomer and Hammill 1997). The subtest measures the child's ability to comprehend the meaning of sentences with an emphasis on the syntax of a sentence, such as understanding the difference between standing near a child and not standing near a child. The task requires no verbalization; the child must select from three pictures the one that most accurately represents the stimulus sentence. It has been standardized to have a mean of 10 and standard deviation of 3. Its reliability has been evaluated through internal consistency (.75-.86) and test-retest (.81). The test is appropriate for students aged 4-8 years. Administration takes 5 to 10 minutes.

Mathematics assessments

- a. **Woodcock-Johnson Applied Problems:** This is a standardized measure of children's mathematical knowledge (McGrew and Woodcock 2001). It assesses children's ability to solve small numerical and spatial problems presented verbally with accompanying pictures of objects. It has a mean of 100 and a standard deviation of 15. Its reliability has been evaluated

through test-retest (.85-.90). The test is appropriate for students aged 2 years and older. Administration takes 5 to 10 minutes.

- b. **Child Math Assessment-Abbreviated (CMA-A) Composite Score:** The CMA-A Composite Score contains four subscales measuring four aspects of early mathematics development: (1) solving simple addition and subtraction problems involving a single set of objects that is initially visible and then hidden from view, (2) constructing a set of objects equal in number to a given set, (3) recognizing shapes, and (4) copying a repeating pattern using sets of objects that vary in color and identity from the objects in the model pattern. The CMA-A Composite Score contains several items per subscale. Each subscale is scored as fraction of items correctly answered. The CMA-A Composite Score is the average of the subscale scores and ranges from 0 to 1. It was adapted specifically for preschool and kindergarten children for the PCER initiative from a more comprehensive early mathematics measure, the Child Mathematics Assessment, by Klein and Starkey (2002) who were also developers of the *Pre-K Mathematics* curriculum and members of the California/New York research team. The authors found a mean CMA-A score of .35 (SE = .05) for children from low-income backgrounds and a mean score of .62 (SE = .04) for children from middle-class backgrounds.
- c. **Building Blocks, Shape Composition task:** This one-item task was adapted for preschool and kindergarten children for the PCER initiative from the Building Blocks assessment tool, which was developed by Clements, Sarama, and Liu (in press). Children are presented with a puzzle shape and a set of pattern blocks. They are asked to use the blocks to fill in the puzzle. The measure is scored on a 0-3 scale with:
- 0 = A student places no shapes *or* places shapes but none “fit”.
 - 1 = A student places shapes with more than 0 percent fitting but with either less than 50 percent fitting or more than two gaps left in the pattern.
 - 2 = A student places shapes with 50 percent or more fitting but leaves one to two gaps or hangovers.
 - 3 = A student places all shapes with no gaps or hangovers.

Teacher report of child behavior

Teacher reports provided the student-level behavior measures used in the evaluation. Preschool teachers gave pre-intervention ratings of child behaviors in the fall of 2003 (after at least a month of class) and post-intervention ratings in the spring of 2004. They rated each child’s behavior (social competence, behavior problems, and classroom performance) using three scales: the Social Skills Rating System (SSRS) Social Skills scale, the SSRS Problem Behaviors scale (Gresham and Elliott 1990), and the Preschool Learning Behaviors Scale (PLBS) (McDermott et al. 2000). Kindergarten teachers provided a longer-term post-intervention rating on the students’ behavior in the spring of 2005 using the two SSRS scales and the Learning Behaviors Scale (LBS) (McDermott et al. 2000). The behavior measures include:

- a. **Social Skills Rating System (SSRS) Social Skills scale:** This standardized measure assesses children’s social competence and problem behaviors. There are three subtests that make up the Social Skills scale: Cooperation, Assertion, and Self-Control. The Social Skills scale is standardized to have a mean of 100 and a standard deviation of 15. The SSRS was developed in two forms—one for children 3 years to 4 years and 11 months old, and the other for students in kindergarten through sixth grade. Its reliability has been evaluated through internal consistency (.93-.94) and test-retest (.85). Administration takes 15 to 25 minutes.
- b. **Social Skills Rating System (SSRS) Problem Behaviors scale:** There are two subtests that make up the Problem Behaviors scale: Externalizing and Internalizing. The Problem Behaviors

scale is standardized to have a mean of 100 and a standard deviation of 15. Higher scores on this scale are indicative of more problem behaviors. Similar to the Social Skills scale in formats and administration, its reliability has been evaluated through internal consistency (.82-.86) and test-retest (.84).

- c. **Preschool Learning Behaviors Scale (PLBS):** The PLBS is a standardized measure of children's behaviors related to classroom learning designed for preschool-age children (McDermott et al. 2000; McDermott, Leigh, and Perry 2002). It is a downward extension of the Learning Behaviors scale. There are four subscales: Confidence/Motivation, Persistence/Attention, Attitude toward Learning, Strategy/Flexibility. The Strategy/Flexibility score is regarded as "experimental" because the dimension was not found to be reliable in the national standardization study, although it was reliable for a Head Start sample. The measure is standardized to have a mean of 50 and standard deviation of 10. Its reliability has been evaluated through internal consistency (.82-.89) and test-retest after 3 weeks (.89). The PLBS is for use with children aged 3 to 5.5 years. Administration time is about 10 minutes.
- d. **Learning Behaviors Scale (LBS), Kindergarten:** The LBS is a standardized measure of children's behaviors related to classroom learning (McDermott et al. 1999). There are four subscales: Confidence/Motivation, Persistence/Attention, Attitude toward Learning, Strategy/Flexibility. The measure is standardized to have a mean of 50 and standard deviation of 10. Its reliability has been evaluated through internal consistency of its four subscales (.82-.92) and test-retest after 2 weeks (.89). The PLBS is for use with children aged 5-17 years. Administration time is 10 minutes.

Classroom observation

Two pre-intervention classroom measures and three post-intervention classroom measures were gathered from preschool classroom observations and used in the evaluation of the curricula. A fourth measure, the Assessment Profile (Abbott-Shim and Sibley 2001), was used but not analyzed because of concerns with the validity of the data collected. Four hours were required to carry out the observation of a preschool classroom using the measures. No observations were made of kindergarten classrooms. Three scales designed to characterize the quality and organization of the classroom and the nature of the interaction between children and the teacher were used in the observations. The Early Childhood Environment Rating Scale-Revised (ECERS-R) (Harms, Clifford, and Cryer 1998) provided an overall measure of the quality of the classroom. The Arnett Caregiver Interaction Scale (Arnett) (Arnett 1989) measured teacher-child interaction on four scales: Positive Interaction, Harshness, Detachment, and Permissiveness. The pre-intervention observation using the ECERS-R and Arnett Scale was conducted in the fall of 2003 and the post-intervention observation in the spring of 2004. The Teacher Behavior Rating Scale (TBRS) (Landry et al. 2002) was added as a post-intervention measure to the spring 2004 observation to capture preschool instructional practices. The TBRS includes scales for teacher instructional practices regarding written expression, print and letter knowledge, phonological awareness, book reading, oral language use, and mathematics concepts.

- a. **Early Childhood Environment Rating Scale-Revised (ECERS-R):** The ECERS-R is a standardized global rating of classroom quality and environment based on the use of space, materials, and experiences to enhance children's development, the daily schedule, and supervision (Harms, Clifford, and Cryer 1998). Data were collected and combined from six of its subscales: Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, and Program Structure. Each sub-scale is rated on a seven-point scale, with higher scores indicating higher quality. The overall ECERS-R score is an average of the scores from each subscale creating a range from 1 to 7. The ECERS-R is based on the original ECERS, developed for preschool classrooms, that has been evaluated for its reliability and predictive validity. Reliability is .92 and inter-rater agreement at the item level is 48 percent exact match

and 71 percent within one point match. Administration requires 140 minutes of a trained classroom observer's time.

- b. **Caregiver Interaction Scale (Arnett):** This is a measure of the quality of the teacher's/caregiver's interaction with a preschool child that includes four scales: Positive Interaction, Harshness, Detachment, and Permissive (Arnett 1989). Each is measured on a four-point scale; higher scores on each scale indicate higher frequency of the associated observed behaviors. For Positive Interaction, higher scores are more optimal; for the remaining three scores, lower scores are more optimal. Administration requires 45 minutes of a trained classroom observer's time. Inter-rater reliability is .80.
- c. **Teacher Behavior Rating Scale (TBRS):** The TBRS is designed to assess specific types of teacher instructional practices that occur in early childhood classrooms (Landry et al. 2002). It was developed by the Center for Improving the Readiness of Children for Learning and Education (CIRCLE) program at the University of Texas Health Sciences Center (this Center was involved in the implementation of two curricula under the PCER study). For the PCER study, the TBRS measures the quantity and quality of the teacher's instructional practices using its Book Reading, Oral Language Use, Phonological Awareness, Print and Letter Knowledge, Written Expression, and Math Concepts subscales. The TBRS was adapted for use in the PCER study and was found to have an inter-rater reliability of .73 in a subset of six classrooms taking part in the study in spring 2004.

Teacher interview and questionnaire

Preschool teachers were interviewed regarding the types and frequency of classroom activities and pedagogy, general classroom information, clarification of observational data, teacher attitudes and beliefs, and teacher background information such as demographics, education and teaching experience and qualifications. Many of the items used were drawn from the Head Start's Family and Child Experiences Survey (Administration for Children and Families 2002a and 2002b) and the National Center for Education Statistics' Early Childhood Longitudinal Study-Kindergarten Cohort (West, Denton, Germino-Hausken 2000). The background information was used to construct covariates for the analysis of the data. Instead of an interview, kindergarten teachers completed a questionnaire that addressed their background, views on kindergarten readiness, classroom resources and activities, instructional practices, and interactions with parents.

Parent interview

Parents were interviewed regarding parent and child demographic information, their own and their child's health and disability status, their assessment of the child's accomplishments and social skills, family-child activities, parenting practices, parental depression, parent involvement with school, and the use of child care. The interview used items from the Head Start's Family and Child Experiences Survey (U.S. Department of Health and Human Services 2002) supplemented with additional measures (Bradley and Caldwell 1984; Gresham and Elliott 1990; Mason and Stewart 1989; Mariner, Zaslow, and Sugland 1998; Radloff 1977). Much of the parent data were collected for descriptive purposes and the demographic information and disability status were used to construct covariates for the analysis of the data. The average length of the parent interview was 94.14 (SD = 25.93) minutes in the fall of 2003 and 105.65 (SD = 47.91) minutes in the spring of 2004.

Study Implementation

The evaluation of the curricula occurred over 2 years, beginning with the preschool year in 2003-04 and continuing through the kindergarten year in 2004-05. The key implementation events in the evaluation of each curricula included randomization of classrooms or programs, consent gathering, teacher training in the use of a treatment curriculum, implementation of the curriculum in the classroom, training the assessors, and

collection of the baseline student and classroom measures and the post-intervention measures in preschool and kindergarten.

Timeline of Implementation

Because research teams independently implemented the curricula and because the schools followed different calendars, the dates and sometimes the order of these events differed between teams and sites within teams. In addition, as RTI and MPR played slightly different roles with their teams, the order of the events also differed by evaluator. Figures 1.1 and 1.2 graphically display the preschool year timeline for the randomization, consent process, start of curriculum implementation, assessor training, and baseline child assessment for the research teams working with RTI and MPR, respectively.

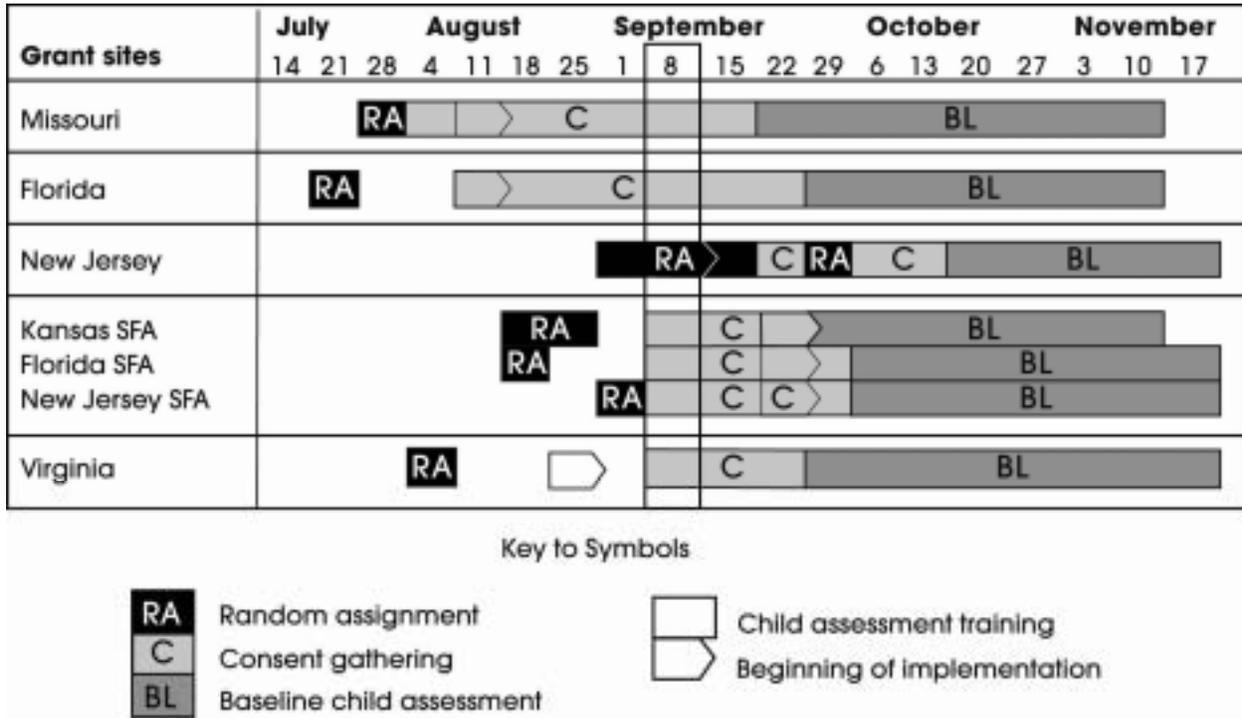
Randomization for the seven teams working with RTI occurred in the pilot year (starting in the fall of 2002). The research teams carried out random assignment. RTI served as the evaluation contractor for the pilot study, collecting all but the parent interview data and analyzing all the data. The pilot year was used to inform the evaluation study protocol and revise the child assessment.

For the preschool evaluation year (2003-04), the pilot-year randomization was carried over for the teams working with RTI but, as noted earlier, teacher turnover and changes in classes required some re-randomization. In all cases new samples of children and parents were recruited for the study. The five teams working with MPR had no pilot year. MPR carried out the randomization in four of the five sites from July through September of 2003. The FSU research team conducted random assignment at their research site.

The consent process followed randomization except for two teams where it occurred concurrently. The start of implementation of the curricula in the classroom ranged from August through October 2003 primarily before baseline data collection began. Although the research teams attempted to collect baseline data close to the beginning of school to avoid student exposure to the treatment curricula before pretesting, there were cases with a lag between the start of implementation and the collection of baseline data ranging from 8 to 49 days (appendix A discusses additional analyses to adjust for possible early treatment effects that might result from these cases). Baseline data collection followed the consent process for the teams working with MPR and ran concurrently for the teams working with RTI. Baseline data collection took 6 to 8 weeks between September and November 2003. Assessors were trained the week of August 4, 2003, for the teams working with RTI and the week of September 8, 2003, for the teams working with MPR.

Pre-kindergarten post-test data were collected in the spring from April to June 2004, depending on school calendars. Student assessments, teacher interviews, teacher reports on behavior, and classroom observations were completed over a 6- to 8-week period. Parent interviews were completed over a 12-week period. Kindergarten post-test data (student assessments, teacher reports, teacher surveys, and parent interviews but no classroom observations) were collected in the spring and summer of 2005 between March and July.

Figure 1.2. Timeline for teams working with Mathematica Policy Research, Inc.



NOTE: The Success for All (SFA) Foundation research team recruited preschool programs in three different programs in three different states (Florida, Kansas, and New Jersey)
 SOURCE: The Preschool Curriculum Evaluation Research (PCER) study.

Child assessors

Training included lectures, live and videotaped demonstrations, paired practices, and discussions. IES staff attended training sessions and observed all training components. Deviations from the training protocol were resolved during the training sessions and in follow-up discussions.

RTI recruited 53 assessors to work on the preschool, fall 2003, baseline data collection. For each research team, one assessor was assigned as the “Lead Assessor,” who was responsible for scheduling and supervising the other assessors and communicating with RTI. In August 2003 RTI conducted a 3-day centralized training for all child assessment staff. Lead assessors received an additional half-day of training. RTI trainers certified all Lead Assessors at the end of training. The Lead Assessors later certified any assessor who did not receive certification at the training. MPR recruited and trained 27 assessors for the fall 2003 data collection. A leader was assigned to coordinate the work of each local assessment team. MPR conducted fall assessor training in September 2003. Twenty-five of the 27 assessors were certified to conduct assessments during the baseline data collection.

For the preschool spring 2004 post-intervention data collection, RTI conducted a 3-day centralized training session in March 2004. Fifty assessors (8 new and 42 returning assessors) were trained and certified to administer the child assessments. MPR staff held a 2-day assessor training in March 2004. Twenty-three new assessors and 15 experienced assessors were trained and certified to complete child assessments.

Table 1.7. Training and support of treatment teachers

Curriculum	Initial training	Ongoing support
<i>Bright Beginnings and Creative Curriculum</i> (Vanderbilt)	<ul style="list-style-type: none"> • 2.5 days at the beginning of the school year 	<ul style="list-style-type: none"> • Onsite consultation four times during the school year
<i>Creative Curriculum</i> (University of North Carolina at Charlotte)	<ul style="list-style-type: none"> • 4 days in August 2002 • Evaluation year refresher sessions (half or full day) Four sessions in NC Five sessions in GA 	<ul style="list-style-type: none"> • Pilot Year: 3 days from 9/02 to 1/03 One small group training session 9 days of technical assistance • Ongoing technical assistance during Evaluation year
<i>Creative Curriculum with Ladders to Literacy</i> (University of New Hampshire)	<ul style="list-style-type: none"> • Minimum of 1 day of <i>Creative Curriculum</i> training • <i>Ladders to Literacy</i> training in September of the preschool year 	<ul style="list-style-type: none"> • Monthly <i>Ladders to Literacy</i> training throughout the school year
<i>Curiosity Corner</i> (Success for All Foundation)	<ul style="list-style-type: none"> • 2 days (12 hours) 	<ul style="list-style-type: none"> • 3 days of follow-up support per teacher
<i>Doors to Discovery and Let's Begin with the Letter People</i> (University of Texas Health Science Center at Houston)	<ul style="list-style-type: none"> • 2 days at the beginning of the pilot school year • 3 days of refresher training in the evaluation year 	<ul style="list-style-type: none"> • 1 day of follow-up training in the pilot year
<i>Early Literacy and Learning Model</i> (University of North Florida)	<ul style="list-style-type: none"> • A 2-day summer training session 	<ul style="list-style-type: none"> • Weekly classroom visits by <i>ELLM</i> literacy coaches • Monthly program-specific literacy team meetings • Quarterly teacher get-togethers
<i>Language-Focused Curriculum</i> (University of Virginia)	<ul style="list-style-type: none"> • 3-day workshop Makeup session for two teachers 	<ul style="list-style-type: none"> • 2 hours in November 2003 • 3 hours in January/February 2004
<i>Literacy Express and DLM Early Childhood Express with Open Court Reading Pre-K</i> (Florida State University)	<ul style="list-style-type: none"> • 4 days for <i>Literacy Express</i> • 6 days for <i>DLM Early Childhood Express with Open Court Reading Pre-K</i> 	<ul style="list-style-type: none"> • Monthly 2-hour professional development meetings • Mentoring visits for half the teachers
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i> (UC-Berkeley and University at Buffalo, SUNY)	<ul style="list-style-type: none"> • 4 days in summer of pilot year • 2-day refresher training in the evaluation year 	<ul style="list-style-type: none"> • 4 days in winter of pilot year • Twice a month training during pilot school year
<i>Project Approach</i> (Purdue and University of WI-Milwaukee)	<ul style="list-style-type: none"> • 3-day workshop 	<ul style="list-style-type: none"> • Mentoring visits • Two 1-day workshops during the school year
<i>Project Construct</i> (University of Missouri-Columbia)	<ul style="list-style-type: none"> • 12-hour module trainings in August (Module 1), October (Module 2), and November (Module 3) 2003 	<ul style="list-style-type: none"> • Four 4-hour on-site consultations • Two 3-hour follow-up workshops
<i>Ready, Set, Leap!</i> (UC-Berkeley)	<ul style="list-style-type: none"> • 4 days spread across the school year (September, November, January, and March) 	<ul style="list-style-type: none"> • Three coaching visits during the school year

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

For the kindergarten spring 2005 post-intervention data collection, RTI trained 58 assessors in March 2005 during a 2-day training session with an additional half-day of training provided for new lead assessors. MPR conducted a 2-day training for all child assessment staff certifying 37 assessors to conduct the kindergarten child assessments.

Parent interviewers

In the 2003-04 preschool evaluation year, RTI and MPR used different approaches to carrying out the parent interviews but comparable interviewer training protocols. For the research teams working with RTI, team staff were responsible for the parent interviews. RTI used a “train the trainer” model to ensure that consistent procedures were used across interviews. Each research team sent their lead interviewer to a centralized 2-day training hosted by RTI in August 2003. The training covered methods for conducting the parent interview and for training others to administer the interview. All lead-parent interviewers were certified at the end of training. Lead interviewers were responsible for training the other interviewers from their research team. Interviews were conducted in person or by phone and interviewers filled in questionnaire booklets that were submitted to RTI. For the teams working with MPR, MPR staff interviewed parents solely by telephone using computer-assisted telephone interviewing (CATI). Interviewers participated in a 2-day, 12-hour CATI training session and 16 hours of project-specific training where they learned about the purposes of the study, the planned use of the data, and the specifics of the parent-interview instrument. The use of two parental data collection methods and RTI’s use of staff from teams implementing the interventions were not examined regarding their potential as a source of bias.

For the spring 2004 data collection, RTI and MPR continued using their interview methods and provided additional training for interviewers. RTI conducted a 1-day refresher training session. MPR gave a 1-day refresher for experienced interviewers in March and a 2-day new interviewer training in April.

In the 2004-05 kindergarten evaluation year, RTI switched to having its own staff interview parents by telephone using the CATI system. Initial training was provided in April 2005, with refresher training provided in May 2005. MPR staff continued conducting telephone interviews using CATI. In March 2005, new MPR interviewers received 16 hours of training and experienced ones received an 8-hour refresher course.

Classroom observations and teacher interviews

Both RTI and MPR data collection staff were trained to conduct the classroom observations and teacher interviews that were done for the 2003-04 preschool evaluation year. RTI recruited classroom observers who had a background in early childhood education and previous experience using the Early Childhood Environment Rating Scale-Revised (ECERS-R) measure. RTI trained observers to use the ECERS-R (Harms, Clifford, and Cryer 1998) and the Arnett Caregiver Interaction Scale (Arnett) (Arnett 1989) in a 2-day session in August 2003. Observers also participated in 2 additional practice days to increase reliability of observations. Observers with limited observation experience participated in 2 additional days of practice in classroom settings. MPR recruited and trained members of their child assessment team to conduct classroom observations. Training was held in September 2003 and included 2 days of classroom training, 1 day of practice observations in the field, and 1 day for a certification visit in the field. Staff were trained to use the ECERS-R and the Arnett Scale.

For the 2003-04 preschool, post-intervention data collection, the Teacher Behavior Rating Scale (TBRS) (Landry et al. 2002) was added to the classroom observation protocol along with the ECERS-R and the Arnett Scale. University of Texas-Houston Health Sciences researchers used a “train the trainer” model to train MPR and RTI staff to conduct classroom observations using the TBRS. RTI staff then conducted a 2-day training session in March 2004. The session reviewed the measures used in the fall of 2003 and taught the use of the TBRS. The RTI-trained observers then spent 2 additional days conducting practice observations. In a 4-day session in March 2004, MPR provided both a refresher training for the ECERS-R and the Arnett Scale plus new training on the TBRS.

RTI calculated inter-rater reliability for the three classroom-observation measures using the rate of agreement between pairs of observers in both the fall and spring pre-kindergarten data collection (table 1.8). Inter-observer reliability data were collected by conducting paired classroom observations at a subset of the research sites. For fall 2003, the data were taken from 11 classrooms from all teams except the North Carolina research team and the Purdue/Wisconsin research team. For spring 2004, the data were taken from six classrooms across all teams except the New Hampshire team and the California/New York research team. Where a team was implementing two curricula, the results included both. The spring TBRS had lower reliability (73%) than the ECERS-R (86% and 96% in the fall and spring) and the Arnett (92% and 96%). MPR did not similarly calculate inter-rater reliability.

Table 1.8. Inter-pair agreement on classroom observations among research teams working with RTI International (RTI), fall 2003 and spring 2004

Research team	ECERS-R (%)	Arnett (%)	TBRS (%)
Fall 2003			
Vanderbilt	88	100	†
University of New Hampshire	80	92	†
University of Texas Health Science Center at Houston	83	88	†
University of Texas Health Science Center at Houston	87	88	†
University of Texas Health Science Center at Houston	83	88	†
University of North Florida ¹	91	92	†
University of North Florida ¹	86	88	†
University of North Florida ¹	86	88	†
UC-Berkeley and University at Buffalo, SUNY ¹	86	92	†
Fall 2003 overall average	86	92	†
Spring 2004			
Vanderbilt	92	100	94
University of North Carolina at Charlotte	97	92	58
University of Texas Health Science Center at Houston	81	96	70
University of Texas Health Science Center at Houston	86	100	81
University of North Florida ¹	100	—	—
Purdue University and University of WI-Milwaukee	100	92	61
Spring 2004 overall average	93	96	73

— Not available.

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

¹ Calculations based on pair observation within a specific subset of schools defined by a local geographic location.

NOTE: ECERS-R: Early Childhood Environment Rating Scale-Revised

Arnett: Caregiver Interaction Scale

TBRS: Teacher Behavior Rating Scale

TBRS only given in spring 2004. For teams using multiple pairs of raters, inter-pair agreement is reported for each pair.

For these teams, the inter-pair agreement was based on a subset of schools.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Response Rates, Attrition, and Mobility

The response rates and the attrition for the child assessments, teacher child reports, and parent interviews are displayed by research team and overall in table 1.9. The baseline data were collected in the fall of 2003 from the original sample with an average response rate of 98 percent for the child assessments, 97 percent for the teacher reports, and 84 percent for the parent interviews. For the first follow-up data collection in the spring of 2004, attrition reduced the percentage of children for whom data were collected to 93 percent of students completing the child assessments, 90 percent having a teacher report, and 79 percent having a parent interview. Further attrition led to an additional decline in the second follow-up data collection in the spring of 2005, with 85 percent of the original sample completing the child assessments, 72 percent having a teacher report, and 75 percent having a parent interview.

Table 1.9. Response rates

Research team	Response rate Fall 2003	Percent of sample with data Spring 2004	Percent of sample with data Spring 2005
Vanderbilt (n = 309)			
Child Assessments	100	94	97
Teacher Report	100	90	90
Parents Interview	82	81	75
UNC-Charlotte (n = 194)			
Child Assessments	98	88	85
Teacher Report	100	88	56
Parents Interview	87	69	71
University of New Hampshire (n = 123)			
Child Assessments	100	85	66
Teacher Report	99	81	50
Parents Interview	16	45	51
Success for All (n = 215)			
Child Assessments	98	95	90
Teacher Report	97	95	82
Parents Interview	91	94	86
University of Texas-Houston (n = 297)			
Child Assessments	99	94	79
Teacher Report	97	86	57
Parents Interview	80	74	68
University of North Florida (n = 244)			
Child Assessments	100	92	89
Teacher Report	96	89	64
Parents Interview	84	81	73
University of Virginia (n = 195)			
Child Assessments	85	96	97
Teacher Report	87	93	81
Parents Interview	93	87	89
Florida State University (n = 297)			
Child Assessments	95	96	80
Teacher Report	96	93	80
Parents Interview	91	84	75

See notes at end of table.

Table 1.9. Response rates—Continued

Research team	Response rate	Percent of sample with data	Percent of sample with data
	Fall 2003	Spring 2004	Spring 2005
UC-Berkeley and University at Buffalo, SUNY (n = 316)			
Child Assessments	99	94	90
Teacher Report	99	94	74
Parents Interview	83	90	78
Purdue and University of WI-Milwaukee (n = 204)			
Child Assessments	100	94	85
Teacher Report	100	90	66
Parents Interview	86	76	70
University of Missouri-Columbia (n = 231)			
Child Assessments	99	90	81
Teacher Report	98	81	68
Parents Interview	92	84	84
UC-Berkeley (n = 286)			
Child Assessments	96	92	87
Teacher Report	96	95	84
Parents Interview	91	82	76
All Teams (n = 2,911)			
Child Assessments	98	93	85
Teacher Report	97	90	72
Parents Interview	84	79	75

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Overall, 15 percent of all the students sampled (426 students) were not included in the analyses: 2 percent non-responders during baseline data collection, and 13 percent through later attrition. For the individual research teams, the percentage of students sampled who were not included in the analysis ranged from 3 percent to 34 percent. There was no evidence of differential sample attrition across the treatment and control groups at each research site (see appendix B).

Child and teacher mobility was part of the reason for attrition. Two hundred and forty-five students moved between the fall of 2003 and the spring of 2004. Of these, 75 remained in the study and 170 attrited. Five classes were not included in the analyses for various reasons (e.g., loss of teacher or teacher's consent, or combining of classes). The students of two of these classes joined other classes in the study with the same condition and so remained in the study. Students in the other three classes attrited from the study. Teacher turnover led to the replacement of 32 teachers (out of 315) during the preschool evaluation year and their students were retained in the analyses.

Contamination

For research teams using school-level random assignment (3 of the 12 teams), the treatment and control groups were in different schools. All of the preschool classrooms in each school were assigned to the treatment or the control condition. Consequently, the risk of contamination from teachers in different conditions exchanging information or materials was minimal.

For research teams using classroom-level random assignment (9 of the 12 teams) to the treatment and control group, the research teams monitored treatment and control classrooms to ensure that they were not sharing any materials or activities with the control group teachers. Based on a review of their classroom observation data and other documentation obtained from teachers, the teams concluded that there was little or no evidence of contamination. The only identified case of contamination concerned a classroom that contained some students who had received a treatment curriculum during the pilot year. This class was dropped from the evaluation of that curriculum.

Fidelity of Implementation

The research teams collected data on the fidelity of implementation for the treatment and control curricula using both a team specific measure and a global implementation rating that can be used for between-curricula comparisons. The global ratings used a four point scale of 0-3 representing Not at All, Low, Medium, or High implementation. The values for multiple classrooms were averaged into a single rating for each curriculum for each research team.

For the treatment curricula, the fidelity ratings ranged from about 1.7 to 2.5 centering around 2 (Medium). The fidelity of the control curricula ranged from 1.5 to 2. The fidelity of implementation for both the treatment and control curricula was rated as Medium.

Sample Description

Tables 1.10-1.13 provide information on all the students, families, teachers, and classrooms involved in the individual studies. Because the PCER study is a set of independent evaluations rather than one single evaluation, no comparison is made between the treatment condition and the control condition aggregated across all sites. Such comparisons are made for each evaluation's treatment and control groups in chapters 2-13. The data presented here are provided to allow comparisons with the sample for each research team described in those chapters.

On average, the students were age 4.6 years at the time of the baseline data collection in the fall of 2003 and age 6.1 years at the time of the kindergarten follow-up in the spring of 2005 (table 1.10). Approximately half (51%) of the children were male. One-third were white non-Hispanic, 43 percent were African American, and 16 percent Hispanic. Less than 7 percent had a disability.

On average, the students' primary caregivers, most often their biological or adoptive mother, were age 32 years at the time of the fall 2003 data collection (table 1.10). Less than half (47%) were married and one-third were never married. Less than half attended or graduated from college (48%), one-third had a high school diploma or GED, and 19 percent did not complete high school. Half were employed full-time, 14 percent part-time, and 34 percent were unemployed.

Almost all the preschool teachers were female (98%) and the majority were White (54%), with one-third African-American (table 1.11). Two-thirds had at least a college degree. On average, they had 12 years of teaching experience and 8 years of experience teaching in pre-kindergarten settings. A majority (87%) of the preschool programs in which they taught were full-day programs (table 1.12). More than half (58%) were public pre-kindergartens, 31 percent were Head Start classes, and child care settings made up the remainder (12%). On average, teachers taught 15 students and the child-staff ratio averaged 7.5 children per teacher.

The kindergarten teachers were also mostly female (98%) and White (74%) with 17 percent African-American (table 1.11). Almost all had at least a BA (97%) with 39 percent having a graduate degree. They averaged 15 years of teaching experience with an average of 9 years of teaching kindergarten. Ninety-three percent of the

Table 1.10. Characteristics of children and parents

Characteristic	Children		Parent	
Age at baseline (years), mean	4.6	(n = 2,845)	31.5	(n = 2,399)
Age at the kindergarten follow-up (years), mean	6.1	(n = 2,480)	†	
Percent male	50.8	(n = 2,900)	†	
Race/ethnicity (%)		(n = 2,636)		(n = 2,410)
White, non-Hispanic	33.5		37.0	
African American, non-Hispanic	43.1		43.4	
Hispanic	15.6		13.8	
Asian or Pacific Islander	‡		1.5	
Native American	‡		0.6	
Multiple/Other	6.1		3.7	
Child disability status (%)	6.5	(n = 2,401)		
Marital status (%)				(n = 2,628)
Married	†		46.7	
Separated/divorced	†		17.1	
Widowed	†		1.2	
Never married	†		35.1	
Educational level (%)				(n = 2,409)
Did not finish high school	†		18.8	
High school diploma or GED	†		32.8	
Some college	†		33.6	
College graduate	†		14.8	
Employment (%)				(n = 2,630)
Full-time	†		50.0	
Part-time	†		14.1	
Unemployed	†		34.0	
Other	†		2.0	

† Not applicable.

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

kindergarten classrooms were full-day and 92 percent of the students were enrolled in public schools (table 1.13). The average number of students per classroom was 20 children. Thirty-nine percent were enrolled in schools where more than 75 percent of the students were eligible for free or reduced-price lunch.

Analysis

Each curriculum was analyzed separately due to the independence of the research teams, the nonrandom assignment of curricula to research teams and sites, and the difference in control conditions among the teams. Because students were nested in classrooms or programs and repeatedly assessed with multiple measures, multi-level models containing student, teacher, and classroom-level covariates were used to address the cross-level correlated errors, allowing for a mixture of random and fixed effects (see appendix B for details). For each curriculum, these models were used to estimate differences in treatment and control group means for each of the 27 measures. The type of model used to analyze each measure depended on the number of time points it was observed.

Table 1.11. Characteristics of preschool and kindergarten teachers

Characteristics	Preschool	Kindergarten
Percent female	98.0	98.0
Race/ethnicity (%)	(n = 313)	(n = 1,085)
White, non-Hispanic	54.3	73.6
African American, non-Hispanic	32.3	17.3
Hispanic	7.0	6.3
Asian or Pacific Islander	‡	1.6
Native American	‡	‡
Multiple/Other	4.5	1.1
Educational level (%)	(n = 315)	(n = 1,088)
Did not finish high school	‡	0.0
High school diploma or GED	19.7	0.9
Associate degree	13.3	0.6
Bachelor of Arts (BA)	46.0	37.4
Post BA	20.0	
Some graduate school	—	21.1
Graduate degree	—	38.8
Other	—	1.3
Preschool teaching credential (%)		
Have a state-awarded preschool certificate	46.4	†
Have a current teaching certificate/license	63.9	†
Have a Child Development Association certificate	27.3	†
Have no credentials	13.7	†
Kindergarten teaching credential (%)		
None	†	‡
Temporary	†	7.8
Alternative	†	2.9
Regular	†	27.1
Highest	†	60.3
Teaching Experience		
Number of years teaching	12.4	15.1
Number of years teaching preschool	8.0	†
Number of years teaching kindergarten	8.0	8.5

— Not available.

† Not applicable.

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004); PCER Kindergarten Teacher Survey (Spring 2005).

Table 1.12. Characteristics of preschools

Characteristics	Percent or average
Full-day programs (% yes)	86.7
Type of school (%)	
Head Start	30.8
Public pre-kindergarten	57.5
Child care	11.8
Average number of children per classroom	15.4
Average teacher-child staff ratio	7.5

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004); PCER Preschool Classroom Observation Battery (Fall 2003 and Spring 2004).

Table 1.13. Characteristics of kindergartens

Characteristics	Percent or average
Type of school (%)	
Public	91.5
Catholic	2.8
Private school, religious	1.5
Private school, non-religious	1.1
Charter	3.1
Type of kindergarten class (%)	
Kindergarten	96.4
Transitional kindergarten	‡
Multigrade/ungraded	2.8
None (child in pre-kindergarten)	‡
Full-day class (% yes)	92.6
Average number of children per classroom, mean	19.7
Average number of teachers and assistants, mean	1.2
Percent of students at a school eligible for free/reduced price lunch	
Schools where less than 25% of the school population is eligible for free/reduced price lunch	26.8
Schools where 26-50% of the school population is eligible for free/reduced price lunch	17.2
Schools where 51-75% of the school population is eligible for free/reduced price lunch	17.7
Schools where more than 75% of the school population is eligible for free/reduced price lunch	39.1

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Kindergarten Teacher Survey (Spring 2005).

Two types of models for repeated measures (spline and simple) were used for outcome measures with comparable data from three or two time points. Analysis of covariance (ANCOVA) was conducted for measures observed at one time point. The more observations of a measure from different time points included in a model, the better able the model is to identify the parameters of interest, in this case the treatment and control group means of the measures. For this reason, the spline repeated measures model is the preferred model followed by the simple repeated measures model and then the ANCOVA. The analysis of each measure uses the most preferred model that can be used given the number of time points the measure was observed. Table 1.14 lists the model used with each measure. The spline repeated measures model was used with the eight student-level academic measures (for the reading, language, and mathematics outcomes) observed three times. The simple repeated measures model was used with four student-level measures (for the preschool phonological awareness and behavior outcomes) and five class-level measures (for the outcomes of classroom quality and teacher-child interaction). The ANCOVA model was used with four student-level measures (for the kindergarten phonological awareness and behavior outcomes) and six classroom-level measures (for the four instruction outcomes). Appendix D contains the covariate-adjusted mean differences and standard errors estimated for each measure using the preferred model.

The repeated measures spline model was used to compare the treatment and control group means at the spring pre-kindergarten and at the spring kindergarten time points for the eight measures that had been observed three times. In addition to a set of covariates, this model included two time variables: (1) the time between the start of the intervention and the spring preschool data collection, and (2) the time between the spring preschool data collection and the spring kindergarten data collection. The model was also used to examine three secondary issues. First, it was used to determine whether there were differences in the treatment and control means at the baseline data collection. Second, for those research teams in which

Table 1.14. Model used with each outcome measure

Outcome	Measure	Times observed	Model
Reading	TERA	3	Spline repeated measures
	WJ Letter Word Identification	3	Spline repeated measures
	WJ Spelling	3	Spline repeated measures
Phonological awareness ¹	Pre-CTOPPP	2	Repeated measures
	CTOPP	1	ANCOVA w/ Pre-K baseline
Language	PPVT	3	Spline repeated measures
	TOLD	3	Spline repeated measures
Mathematics	WJ Applied Problems	3	Spline repeated measures
	CMA-A	3	Spline repeated measures
	Shape Composition ²	3	Spline repeated measures
Pre-kindergarten behavior ¹	SSRS Social Skills	2	Repeated measures
	SSRS Problem Behavior	2	Repeated measures
	PLBS	2	Repeated measures
Kindergarten behavior ¹	SSRS Social Skills	1	ANCOVA w/ Pre-K baseline
	SSRS Problem Behavior	1	ANCOVA w/ Pre-K baseline
	LBS	1	ANCOVA w/ Pre-K baseline
Classroom quality	ECERS-R	2	Repeated measures
Teacher-child interaction	Arnett Detachment	2	Repeated measures
	Arnett Harshness	2	Repeated measures
	Arnett Permissiveness	2	Repeated measures
	Arnett Positive Interaction	2	Repeated measures
Literacy instruction	TBRs Written Expression	1	ANCOVA
	TBRs Print and Letter Knowledge	1	ANCOVA
Phonological instruction	TBRs Phonological Awareness	1	ANCOVA
Language instruction	TBRs Book Reading	1	ANCOVA
	TBRs Oral Language	1	ANCOVA
Mathematics instruction	TBRs Math Concepts	1	ANCOVA

¹ Pre-kindergarten and kindergarten measures not on the same scale

² Building Blocks, Shape Composition task

NOTE: ANCOVA: Analysis of covariance. Refer to the glossary for abbreviations of the measures. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten).

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

curriculum implementation occurred before baseline data collection, the model was used to project backwards from the baseline to the time of implementation to test for group differences at that point (and so address early treatment affects). This was done using the first time variable, which provides the treatment and control group growth rates (slope) during preschool, and an assumption of straight-line growth that allowed the growth rate to be project backward to the start of implementation. Third, the treatment and control group growth rates were compared during preschool (using the first time variable) and kindergarten (using the

second time variable). Chapter 1 discusses the primary results from the comparison of the treatment and control group means at spring pre-kindergarten and at spring kindergarten. The results pertaining to the secondary issues are discussed in appendix A.

For the four student-level measures and five classroom-level measures with observations at two time points, a simple repeated measures model was used to compare the treatment and control group means at spring pre-kindergarten. Along with the set of covariates, this model included one time variable: the time between the start of the intervention and the spring preschool data collection. The model was also used to address the three secondary issues: (1) group mean differences at the baseline, (2) differences at the start of treatment, and (3) difference in rates of growth in pre-kindergarten (but not kindergarten). The primary results from the comparison of the treatment and control group means at spring pre-kindergarten are discussed in chapter 1 and the secondary analysis results in appendix A.

ANCOVA models were used to estimate the difference in mean measures between the treatment and control groups in the spring of pre-kindergarten or kindergarten time points when only one observation was available. The availability of only one observation of a measure occurred in two situations. First, four of the kindergarten student measures (the CTOPP, SSRS Social Skills scale, SSRS Problem Behaviors scale, and LBS) were not on the same scales as the pre-kindergarten measures. The ANCOVA model for these kindergarten measures included students' scores on the respective pre-kindergarten scale as a covariate to address any differences in the groups that occurred despite randomization. Second, six pre-kindergarten classroom instruction measures were based on the TBRS that was given only in the spring pre-kindergarten. Group mean differences for these were estimated using an ANCOVA without a similar baseline covariate. These models may be biased by any initial differences in instruction that occurred despite randomization as there was no baseline measure. Both ANCOVA models included the student, teacher, and classroom covariates used in the repeated measures models.

All three types of models included a set of student and classroom covariates to increase the precision of the estimates by accounting for chance baseline differences between the groups on those characteristics. The child-level covariates were children's age, race/ethnicity, gender, parent report of disability, and mothers' education. The classroom-level covariates included teachers' educational attainment, previous teaching experience, teachers' race/ethnicity, child/adult ratio in the preschool classroom, average class size, and city size. This set of covariates was selected based on preliminary data analyses showing a relationship between these variables and the measures. Another set of covariates (listed in appendix B) was not included because no such relationship was found. Inclusion of the student-age covariate required that the analysis of eight of the student-level academic measures use non-standardized scores because their standardized scores account for developmental growth associated with a student's age. As a result, raw scores were used for the TERA, Pre-CTOPPP, CTOPP, PPVT, and TOLD Grammatical Understanding subtest and the three WJ measures were transformed into W scores using the Rasch ability scale (see appendix B for details).

Results

The goal of the PCER initiative was to identify the impact of the 14 preschool curricula on five student-level outcomes (reading, phonological awareness, language, mathematics, and behavior) and six classroom-level outcomes (classroom quality, teacher-child interaction, and four types of instruction). Each outcome was based on one or more of the measures (table 1.14); therefore, the process of determining a curriculum's impact on the outcomes required two steps. First, the models were used to identify average differences in the 27 measures between the students receiving the treatment curriculum and those receiving the control, and determine their statistical significance. Second, a criterion was applied to a set of measures that made up a specific outcome to determine whether the results for that group of measures showed a finding that the curriculum had an impact on that outcome. This process is described in the following order: (1) the model results for the 27 measures, (2) considerations regarding the efficacy nature of the evaluations, the statistical

power of the analyses, and the use of multiple comparisons, and, (3) the criteria applied to the measures to determine findings for each outcome. The findings are then described in the final section of chapter 1.

Model Results

The models tested the difference between the means of the treatment versus the control group for each measure. Tables 1.15-1.17 display this difference as an effect size and note which differences are statistically significant. Effect sizes provide a relative measure of the magnitude of differences allowing comparisons of the results for the different measures, the different years, and the different models. Cohen's d was used to determine the effect size for each measure: the mean of the control group was subtracted from the mean of the treatment group and the difference was divided by the pooled standard deviation of the treatment and control groups. The pooled standard deviation is specific to each research team and number of time points included in the model (see appendix B for details).

In tables 1.15-1.17, the measures are grouped under their corresponding student-level and classroom-level outcomes. Table 1.15 identifies the impacts of each curriculum on the student-level measures in pre-kindergarten. Ten curricula showed no statistically significant impacts on any of the student-level measures while five showed significant impacts on some measures (three curricula affected only one measure). Table 1.16 identifies the impacts of the curricula on student-level measures in kindergarten. Nine curricula showed no statistically significant impacts on any of the student-level measures in kindergarten and six do (five curricula affected one or two measures). Table 1.17 shows the impacts of the curricula on the preschool classroom-level measures. Seven curricula had no statistically significant impact on these measures and eight curricula showed an impact (five curricula showed an impact on one or two measures).

Considerations: Efficacy, Power, and Multiple Comparisons

The experimental design used to generate the results displayed in tables 1.15-1.17 is a rigorous form of evaluation. The evaluations' focus on the efficacy of the curricula, the statistical power of each evaluation to find an impact, and the need to make multiple comparisons due to the many outcomes should be considered when reviewing the results.

Efficacy trials

The evaluations conducted under the PCER study were efficacy trials—that is, they were intended to determine whether the curricula are effective under specified conditions. Those conditions included public pre-kindergarten programs serving predominantly low-income families in a particular location with ongoing professional development support from researchers. The results from efficacy evaluations have less generalizability than results from evaluations of interventions implemented at scale. The lack of widespread implementation prevents the conclusion that the results broadly apply.

Statistical power

The original IES Request for Applications to which the 12 research teams successfully responded required that each team include a minimum of 10 treatment and control classrooms or preschool programs (half treatment and half control) and 150 students. All teams exceeded the classroom/program requirement. After the data were collected, achieved power was calculated to determine the minimum detectable effect (MDE) sizes (d) for each evaluation. The MDEs calculated using achieved power are lower than if calculated before a study begins as they take into account the smaller actual samples that occur due to non-response and attrition. Table 1.18 displays the MDEs by research team for four composite measures (Reading, Language, Mathematics, and Behavior) that combine the preschool child-level measures under each of these outcomes (Reading also includes the Pre-CTOPPP). Each cell of the table contains a higher more conservative MDE and a lower less conservative one. The MDEs range from .34 to .69 across the composites and teams.

The MDEs were calculated using the following values. The probability of a Type 1 error (α) was set at .05. The number of classrooms or programs for each evaluation was the number of clusters (J) and the number of

students per each was the cluster size (n). The variance (R^2) explained by a covariate (pre-test) was calculated by comparing a model estimating the spring preschool composite using only the pre-test (the composite score in fall of preschool) with the full model. These values ranged from .51 for the Math composite to .67 for the Reading composite. The intraclass correlation (ρ) was set at .05 and .15 (the latter a more conservative value) based on findings from other early childhood evaluations (Schochet 2005). Optimal Design Software (Spybrook et al. 2006) was used to calculate the MDEs.

Multiple comparisons

The analysis of 27 measures (some of which occurred multiple times) required multiple comparisons to be made for each evaluation. The chances of observing a significant finding, when in fact there is not one (Type 1 error), increase with multiple comparisons. If the measures involved are related, the chances increase further. Statistical adjustments for multiple comparisons were not made for the evaluations of the curricula. In part to offset the chance of such error, the findings are not based on the individual measures. Instead, they are based on a set of criteria that define how the results for the measures are translated into conclusions regarding the student and classroom-level outcomes under which the measures are grouped. Where possible, these criteria require that a finding be based on at least two statistically significant measures.

Criteria for Findings

Four of the five student-level outcomes had two to three outcome measures associated with them (phonological awareness only had one per grade), as did three of the six classroom-level outcomes. The measures within an outcome were conceptually related to one another and sufficiently inter-correlated that an effect on one would not be expected to appear, except by chance, without indications of some effect on the others. The following criteria were applied to the results for the measures to determine whether a curriculum had a treatment effect on each student-level outcome for pre-kindergarten and for kindergarten:

- The reading, mathematics, and behavior outcomes each contained three measures. The finding that a curriculum had an effect on any of these three outcomes required at least two of the three measures to have had a statistically significant effect with the same sign and no significant effect with the opposite sign.
- The language outcome contained two measures. A finding of an outcome effect required at least one of the two measures to have had a statistically significant effect and no significant effect with the opposite sign.
- The phonological awareness outcome contained one measure. A finding of an outcome effect required this measure (Pre-CTOPPP in pre-kindergarten and CTOPP in kindergarten) to have had a statistically significant effect.

A similar set of rules was used to determine whether a curriculum had a treatment effect on each pre-kindergarten classroom-level outcome:

- The classroom quality outcome contained one measure. A finding of an outcome effect required this measure to have had a statistically significant effect.
- The teacher-child relationship outcome contained four measures. A finding of an outcome effect required at least two of the four measures to have had a statistically significant effect in the same direction and no statistically significant effects with the opposite direction. For these measures, direction concerns desirability of the effect; a desirable effect was a positive sign for the Positive Interaction scale and a negative effect for the other three scales.
- The early literacy instruction outcome and the early language instruction outcome each contained two measures. A finding of an outcome effect required at least one of the two measures to have had a statistically significant effect and no significant effect with the opposite sign.

- The phonological instruction outcome and the mathematics instruction outcome each contained one measure. A finding of an outcome effect required the measure to have had a statistically significant effect.

Findings

Through the application of the criteria, each curriculum's impact on each outcome was determined. These findings are presented using two forms of organization: findings by outcome and findings by curriculum. Under the Findings by Outcome, those curricula affecting each of the five student-level (for pre-kindergarten and kindergarten) and six classroom-level outcomes (for pre-kindergarten) are identified. Under the Findings by Curriculum, each curriculum is discussed as to its effects on the outcomes.

The findings described are presented in tables 1.19 and 1.20. Table 1.19 shows the impacts of each curriculum on the student-level outcomes for both pre-kindergarten (pre-K) and kindergarten (K). A blank cell stands for no effect, a plus sign (+) means a positive effect, a minus sign (-) means a negative effect, and a zero (0) signifies no effect in one grade when there is an effect in the other. Table 1.20 shows the impact of each curriculum on the classroom-level outcomes using the same symbols.

Findings by Outcome

Two of the 14 intervention curricula had impacts on the student-level outcomes for the pre-kindergarten year (table 1.19). *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* positively affected reading, phonological awareness, and language. *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* curricula positively affected mathematics.

In the kindergarten year, four of the curricula had impacts on the student-level outcomes though three of these did not have impacts during the pre-kindergarten year (table 1.19). *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* continued to have positive effects on reading, phonological awareness, and language in kindergarten as it did in pre-kindergarten. *Curiosity Corner*, which had no effects in pre-kindergarten, was found to positively affect reading in kindergarten. *ELLM*, which had no effects in pre-kindergarten, was found to positively affect language in kindergarten. *Project Approach*, which had no effects in pre-kindergarten, was found to negatively affect behavior in kindergarten.

Eight of the 14 treatment curricula had a positive effect on the pre-kindergarten classroom-level outcomes (table 1.20). *Bright Beginnings* affected early literacy instruction and phonological awareness instruction. *Creative Curriculum* (as implemented by the North Carolina research team but not by the Tennessee research team) affected classroom quality, teacher-child interaction, early literacy instruction and early language instruction. *Creative Curriculum with Ladders to Literacy* affected early literacy instruction. *Curiosity Corner* affected early language instruction. *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* affected phonological awareness instruction. *Doors to Discovery* affected early literacy instruction and early language instruction. *Let's Begin with the Letter People* affected classroom quality and early literacy instruction. *Literacy Express* affected classroom quality and phonological awareness instruction.

Table 1.15. Effect sizes for student-level measures: Pre-kindergarten

Outcome/Measures	Curricula														
	BB	CC (V)	CC (UNC)	CC with Ldrs	Curiosity Corner	DD	LB	ELLM	LFC	DLM with OC	LE	Pre-K Math	PA	PC	RSL
Reading															
TERA	.39*	.02	-.08	-.30	.10	.06	.02	.15	.16	.68***	.17	.13	.14	.00	.08
WJ Letter Word Identification	.35	.16	-.08	-.16	.09	.10	.10	-.05	.11	.51**	.30	-.01	.42	-.05	.01
WJ Spelling	.18	.19	-.18	.30	.04	.06	.17	.11	.25	.46**	.05	.20	.27	-.15	.20
Phonological awareness															
Pre-CTOPPP	-.07	.10	.02	-.16	.18	.18	-.13	.18	.20	.32*	.14	.04	.05	.10	-.09
Language															
PPVT	.13	.23	.08	-.38	-.01	.15	-.03	.17	.02	.40*	.17	.17	.16	.03	.15
TOLD	.09	.07	-.16	-.22	-.08	.17	.08	.15	.01	.40**	-.04	.17	.15	-.05	-.11
Mathematics															
WJ Applied Problems	.16	.17	.20	-.14	.10	.01	-.10	.10	.20	.36**	.05	.22	.07	.06	.04
CMA-A Mathematics Composite	.14	.10	-.10	.18	.01	.13	.15	.01	.08	.17	-.02	.44**	.18	-.11	-.24*
Shape Composition ¹	-.03	.12	.19	.02	.16	-.13	.21	-.14	.08	.24	-.01	.96***	.27	-.42**	.08
Behavior															
SSRS Social Skills	-.27	.03	.05	-.25	-.06	-.18	-.27	-.06	-.42	-.11	-.06	.22	.04	.22	-.05
SSRS Problem Behavior	.23	.07	-.16	-.01	.43	-.14	-.06	-.24	.37	.11	-.31	-.09	.50	-.08	-.03
PLBS	.04	.14	.07	-.08	-.25	-.18	-.44	.14	-.27	-.16	.17	.09	-.31	.00	.07

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

¹ Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures. Abbreviations for the curricula are:

BB: *Bright Beginnings*

CC (V): *Creative Curriculum (Vanderbilt University)*

CC (UNC): *Creative Curriculum (University of North Carolina at Charlotte)*

CC with Ldrs: *Creative Curriculum with Ladders to Literacy*

DD: *Doors to Discovery*

LB: *Let's Begin with the Letter People*

ELLM: *Early Literacy and Learning Model*

LFC: *Language-Focused Curriculum*

DLM with OC: *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

LE: *Literacy Express*

Pre-K Math: *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

PA: *Project Approach*

PC: *Project Construct*

RSL: *Ready, Set, Leap!*

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table 1.16. Effect sizes for student-level measures: Kindergarten

Outcome/Measures	Curricula														
	BB	CC (V)	CC (UNC)	CC with Ldrs	Curiosity Corner	DD	LB	ELLM	LFC	DLM with OC	LE	Pre-K Math	PA	PC	RSL
Reading															
TERA	-.07	.10	-.04	-.54	.43*	-.05	-.13	.30	.05	.76**	-.11	.31	.29	-.03	.01
WJ Letter Word Identification	.09	.38	.00	-.27	.43*	-.09	-.18	.00	.02	.50**	.08	.22	.03	.16	-.12
WJ Spelling	.06	.25	-.05	-.08	.20	-.12	-.06	.04	.11	.22	.06	.03	.14	.00	.04
Phonological awareness															
CTOPP	.01	.06	.06	-.10	.25	-.09	-.13	.08	.03	.38*	.08	-.11	-.17	-.12	-.02
Language															
PPVT	.07	.12	.15	-.30	.14	.18	.00	.34*	-.09	.48**	.16	.11	.10	.10	-.02
TOLD	.16	.11	-.17	-.06	.15	.06	-.12	.44**	-.07	.46**	.10	.08	.32	.01	-.03
Mathematics															
WJ Applied Problems	.13	.17	.09	-.33	.26	-.02	-.13	.26	.11	.48***	-.02	.13	.27	.08	.00
CMA-A Mathematics Composite	.07	.05	.14	-.19	-.05	-.16	-.07	-.05	.00	.13	-.21	.13	.22	-.06	-.10
Shape Composition ¹	.15	.00	-.01	-.10	.32	-.12	-.06	.03	.06	.09	-.14	.41***	.24	.12	.03
Behavior															
SSRS Social Skills	.03	.35	-.12	.17	.32	-.05	.24	.27	-.07	-.18	-.37	.06	-.44*	.12	-.03
SSRS Problem Behavior	.24	-.05	.08	.02	-.08	.46	.06	.23	-.05	.01	.22	-.01	.49*	.07	.07
LBS	.30	.08	-.20	-.11	.11	-.32	-.10	.04	.10	-.13	-.38*	.01	-.42*	-.02	-.01

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

¹ Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures. Abbreviations for the curricula are:

BB: *Bright Beginnings*

CC (V): *Creative Curriculum* (Vanderbilt University)

CC (UNC): *Creative Curriculum* (University of North Carolina at Charlotte)

CC with Ldrs: *Creative Curriculum with Ladders to Literacy*

DD: *Doors to Discovery*

LB: *Let's Begin with the Letter People*

ELLM: *Early Literacy and Learning Model*

LFC: *Language-Focused Curriculum*

DLM with OC: *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

LE: *Literacy Express*

Pre-K Math: *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

PA: *Project Approach*

PC: *Project Construct*

RSL: *Ready, Set, Leap!*

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table 1.17. Effect sizes for classroom-level measures: Pre-kindergarten

Outcome/Measure	Curricula														
	BB	CC (V)	CC (UNC)	CC with Ldrs	Curiosity Corner	DD	LB	ELLM	LFC	DLM with OC	LE	Pre-K Math	PA	PC	RSL
Global classroom quality															
ECERS-R	.80	.45	1.66*	-.71	-.48	.39	.82*	-.48	—	.34	1.29*	.05	-.19	.54	.16
Teacher-child interaction															
Arnett Detachment	.19	-.16	-1.68*	.51	-.41	-.07	-.07	-.41	—	-.06	-1.09	-.37	.57	.12	.19
Arnett Harshness	.12	-.12	-.70	-.26	.14	-.38	-.95*	-.40	—	-.70	-.84	.18	.86	-.13	.30
Arnett Permissiveness	.16	.51	-1.01	1.02	-.98	.13	-.05	-.24	—	.05	.51	-.45	-.43	-.02	-.24
Arnett Positive Interactions	.41	-.15	1.65**	.03	.02	.38	.48	.29	—	.43	.56	.16	-.99	.46	.04
Language instruction															
TBRS Book Reading	1.03	-.47	.28	-.32	2.06**	1.18*	.63	.32	-.79	.01	.49	.07	-.76	.81	-.18
TBRS Oral Language	.39	-.07	1.80**	-.50	.37	.59	.44	.14	.87	-.33	.25	.19	-.42	.52	-.24
Phonological instruction															
TBRS Phonological Awareness	1.53*	1.97	-.10	-.19	.44	.58	.66	.53	.92	1.41*	1.26*	.38	-1.19	.01	.22
Literacy instruction															
TBRS Print and Letter Knowledge	1.51*	1.81	1.02	.75	-.99	.90*	.99*	.41	.33	.91	1.07	.07	.34	.34	-.02
TBRS Written Expression	1.61*	1.99	1.73**	1.13*	-.54	.62	.60	-.22	.99	-.58	-.03	-.12	.62	.43	.10
Mathematics instruction															
TBRS Math Concepts	.98	1.48	.75	.44	-.33	.37	.24	-.92	.20	-.46	-.12	.57	-.64	.53	-.10

— Not available.

* $p < .05$; ** $p < .01$

NOTE: Refer to the glossary for abbreviations of the measures. Abbreviations for the curricula are:

BB: *Bright Beginnings*CC (V): *Creative Curriculum* (Vanderbilt University)CC (UNC): *Creative Curriculum* (University of North Carolina at Charlotte)CC with Ldrs: *Creative Curriculum with Ladders to Literacy*DD: *Doors to Discovery*LB: *Let's Begin with the Letter People*ELLM: *Early Literacy and Learning Model*LFC: *Language-Focused Curriculum*DLM with OC: *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*LE: *Literacy Express*Pre-K Math: *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*PA: *Project Approach*PC: *Project Construct*RSL: *Ready, Set, Leap!*

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table 1.18. Achieved minimum detectable effects on the reading, language, mathematics, and behavior composites of measures

Research team	Reading composite	Language composite	Mathematics composite	Behavior composite
Vanderbilt University	.47 to .56	.48 to .59	.49 to .60	.49 to .59
University of North Carolina at Charlotte	.46 to .52	.47 to .54	.48 to .56	.47 to .55
University of New Hampshire	.58 to .65	.59 to .67	.60 to .69	.59 to .67
Success for All Foundation	.44 to .51	.45 to .53	.46 to .55	.45 to .53
University of Texas Health Science Center at Houston	.43 to .46	.43 to .48	.44 to .49	.43 to .48
University of North Florida	.39 to .43	.40 to .45	.40 to .46	.40 to .45
University of Virginia	.48 to .56	.50 to .59	.50 to .61	.50 to .60
Florida State University	.52 to .63	.54 to .67	.55 to .69	.54 to .67
UC-Berkeley and University at Buffalo, SUNY	.34 to .37	.34 to .38	.35 to .39	.34 to .39
Purdue University and University of WI-Milwaukee	.48 to .57	.50 to .61	.50 to .63	.50 to .61
University of Missouri-Columbia	.42 to .48	.43 to .50	.43 to .51	.43 to .50
UC-Berkeley	.36 to .39	.37 to .41	.37 to .42	.37 to .41

NOTE: $\alpha = .05$; ICC (p) = .05 and .15R²: Reading composite = .67

Language composite = .57

Mathematics composite = .51

Behavior composite = .56

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table 1.19. Findings by student-level outcomes

Curricula	Phonological				Behavior
	Reading	awareness	Language	Mathematics	
<i>Bright Beginnings</i>					
<i>Creative Curriculum (Vanderbilt)</i>					
<i>Creative Curriculum (UNC-Charlotte)</i>					
<i>Creative Curriculum with Ladders to Literacy</i>					
<i>Curiosity Corner</i>		Pre-K: 0 K: +			
<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>	Pre-K: + K: +	Pre-K: + K: +	Pre-K: + K: +		
<i>Doors to Discovery</i>					
<i>Early Literacy and Learning Model</i>			Pre-K: 0 K: +		
<i>Language-Focused Curriculum</i>					
<i>Let's Begin with the Letter People</i>					
<i>Literacy Express</i>					
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>				Pre-K: + K: 0	
<i>Project Approach</i>					Pre-K: 0 K: -
<i>Project Construct</i>					
<i>Ready, Set, Leap!</i>					

NOTE: Abbreviations of the findings are:

Pre-K: Pre-kindergarten

K: Kindergarten

+: Finding of a positive impact

-: Finding of a negative impact

Blank Cell: Finding of no impact

0: Finding of no impact (when an impact is found for the other grade)

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table 1.20. Findings by classroom-level outcomes, pre-kindergarten year only

Curricula	Classroom quality	Teacher-child interaction	Early literacy instruction	Phonological awareness instruction	Early language instruction	Math concepts instruction
<i>Bright Beginnings</i>			+	+		
<i>Creative Curriculum</i> (Vanderbilt)						
<i>Creative Curriculum</i> (UNC-Charlotte)	+	+	+		+	
<i>Creative Curriculum with Ladders to Literacy</i>			+			
<i>Curiosity Corner</i>						+
<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>				+		
<i>Doors to Discovery</i>			+		+	
<i>Early Literacy and Learning Model</i>						
<i>Language-Focused Curriculum</i>						
<i>Let's Begin with the Letter People</i>	+		+			
<i>Literacy Express</i>	+			+		
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>						
<i>Project Approach</i>						
<i>Project Construct</i>						
<i>Ready, Set, Leap!</i>						

NOTE: Abbreviations of the findings are:

+: Finding of a positive impact

Blank Cell: Finding of no impact

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Findings by Curriculum

Each curriculum is discussed separately and cross-curriculum comparisons are not made. The type of pre-kindergarten program involved in the evaluation and the control curricula are described (though the results should not be used to evaluate any control curricula). Impacts on the outcomes are then presented in the following order: (1) student-level outcomes in pre-kindergarten, (2) student-level outcomes in kindergarten, and (3) classroom-level outcomes in pre-kindergarten.

Bright Beginnings

Bright Beginnings and its control were implemented in state pre-kindergarten classrooms in Tennessee. In the control classrooms, teachers used teacher-developed curricula with a focus on basic school readiness. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on early literacy instruction and phonological awareness instruction.

Creative Curriculum—Vanderbilt University

Creative Curriculum and its control were implemented in state pre-kindergarten classrooms in Tennessee. In the control classrooms, teachers used teacher-developed curricula with a focus on basic school readiness. No impacts regarding pre-kindergarten or kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Creative Curriculum—University of North Carolina at Charlotte

Creative Curriculum and its control were implemented in full-day Head Start programs in North Carolina and Georgia. In the control condition, teachers used teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on overall classroom quality, teacher-child relationships, early literacy instruction, and early language instruction.

Creative Curriculum with Ladders to Literacy

Ladders to Literacy was implemented in full-day and half-day Head Start classrooms in New Hampshire as a supplementary curriculum in conjunction with *Creative Curriculum*. In the control condition, teachers used only *Creative Curriculum*. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on early literacy instruction.

Curiosity Corner

Curiosity Corner and its control were implemented in full-day preschool programs in three different states (Florida, Kansas, and New Jersey). In the control condition, teachers used a variety of preschool curricula including the *Creative Curriculum* and *Animated Literacy* curriculum models, and teacher-developed curricula. No impacts regarding pre-kindergarten student-level outcomes were found. A positive impact on reading was found at the end of kindergarten. A positive impact was found at the classroom level on early language instruction.

DLM Early Childhood Express supplemented with Open Court Reading Pre-K

The evaluation of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* took place in public pre-kindergarten classrooms in Florida. In the control condition, teachers were provided with the *High/Scope* curriculum. A positive impact was found on reading, phonological awareness, and language development in both pre-kindergarten and kindergarten. A positive impact was found at the classroom level on phonological awareness instruction.

Doors to Discovery

Doors to Discovery and its control were implemented in full-day Head Start and public pre-kindergarten (Title I and non-Title I) programs in Texas. In the control condition, teachers used teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on early literacy instruction and early language instruction.

Early Literacy and Learning Model (ELLM)

The *Early Literacy and Learning Model (ELLM)* curriculum was implemented in combination with the existing comprehensive curricula that were in use in the control group classrooms in Florida. Several curricula were used in the control classrooms including *Creative Curriculum*, *Beyond Centers and Circletime*, *High Reach*, and *High/Scope*. No impacts regarding pre-kindergarten student-level outcomes were found. A positive impact on language development was found at the end of kindergarten. No impacts were found on the classroom-level outcomes.

Language-Focused Curriculum

The *Language-Focused Curriculum (LFC)* was implemented in full-day Head Start and public pre-kindergarten classrooms in Virginia. The control teachers reported using *High/Scope* curriculum materials. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. No impacts were found on the classroom instruction outcomes. Impacts on classroom quality and teacher-child interaction outcomes could not be determined because of unreliable (inflated) data from eight classrooms on the relevant measures.

Let's Begin with the Letter People

Let's Begin with the Letter People and its control were implemented in full-day Head Start and public pre-kindergarten (Title I and non-Title I) programs in Texas. In the control condition, teachers used teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten or kindergarten student-level outcomes

were found. A positive impact was found at the classroom level on classroom quality and early literacy instruction.

Literacy Express

Literacy Express and its control were implemented in public pre-kindergarten classrooms in Florida. In the control condition, teachers were provided with the *High/Scope* curriculum. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. A positive impact was found at the classroom level on classroom quality and phonological awareness instruction.

Pre-K Mathematics supplemented with DLM Early Childhood Express Math software

The evaluation of *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* took place in Head Start and public pre-kindergarten classrooms in California and New York. Several curricula were used in the control condition including *Creative Curriculum*, *High/Scope*, *Montessori*, specialized literacy curricula, and local school district and teacher-developed curricula. A positive impact was found on student's mathematical knowledge at the end of pre-kindergarten. No impacts on the kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Project Approach

The *Project Approach* curriculum was implemented in public pre-kindergarten classrooms in Wisconsin. In the control classrooms, teachers reported implementing their own teacher-developed, nonspecific curricula. No impacts on the pre-kindergarten student-level outcomes were found. A negative impact on behavior was found at the end of kindergarten. No impacts were found on the classroom-level outcomes.

Project Construct

Project Construct was implemented in full-day child-care centers in Missouri. In the control schools, teacher-developed generic curricula were implemented. No impacts on the pre-kindergarten or kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Ready, Set, Leap!

Ready, Set, Leap! was implemented in pre-kindergarten programs in New Jersey. In the control condition, teachers used the *High/Scope* approach. No impacts on the pre-kindergarten and kindergarten student-level outcomes were found. No impacts were found on the classroom-level outcomes.

Chapter 2. *Bright Beginnings* and *Creative Curriculum*: Vanderbilt University (Tennessee site)

Curriculum

The Vanderbilt University (Tennessee) researchers evaluated *Bright Beginnings* and *Creative Curriculum*.¹

Bright Beginnings

Bright Beginnings is an integrated curriculum with a focus on language and early literacy. It is based in part on the *High/Scope* and *Creative Curriculum* models, with an added focus on literacy skills that are designed to promote school readiness. The curriculum goals are to provide a child-centered, literacy-focused program that is consistent with developmentally appropriate practice and to include instruction that addresses the cognitive, social, emotional, and physical development of young children. The curriculum was especially designed to provide continuity in the pre-kindergarten to second-grade curricula. *Bright Beginnings* includes nine curriculum units that are linked to the program components:

- language and literacy;
- mathematics;
- social and personal development;
- healthful living;
- scientific thinking;
- social studies;
- creative arts;
- physical development; and
- technology.

The classroom environment is designed to encourage children's active exploration and interaction with adults, other children, and classroom materials. Teachers conduct ongoing assessments of children as they engage in a range of classroom activities. The curriculum also includes a Family-School Connection link. Parents sign a parent-school partnership agreement that requires a parent/caregiver to be actively engaged in the child's education.

Creative Curriculum

Creative Curriculum is a comprehensive curriculum for 3- to 5-year-old children. The curriculum addresses four areas of development:

- social/emotional;
- physical;
- cognitive; and
- language development.

¹ The University of North Carolina at Charlotte research team also evaluated *Creative Curriculum*.

Creative Curriculum requires the physical space of the classroom to be structured into 10 interest areas: blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, and computers. Time is also allotted for outdoor activities. The 10 interest areas are designed to address curriculum content, such as literacy, mathematics, science, social studies, the arts, technology, and process skills, such as observing, exploring, and problem solving. *Creative Curriculum* includes a Developmental Checklist teachers are asked to use in ongoing assessments of child progress.

Sample

The Tennessee research team recruited 36 public pre-kindergarten classrooms in seven school districts in six different counties. All of the selected programs were full-day pre-kindergarten programs. Teachers were recruited in July of the preschool year, curriculum training occurred in August, and parental consent was obtained in late August 2003 and early September 2003. A total of 36 teachers/classrooms and 558 parents and children were recruited for participation in the site-level study. A subset of that sample (21 classrooms and teachers) including 309 children and 300 parents (103 in the *Bright Beginnings* treatment group, 101 in the *Creative Curriculum* treatment group, and 105 in the control group) were included in the PCER study sample. Data were collected on 309 children and 252 parents at the time of the fall baseline data collection.

In the follow-up year of the evaluation, the sample of schools went from 19 preschools to 64 schools in kindergarten. The evaluation sample of classrooms went from 21 preschool to 134 kindergarten classrooms. The kindergarten sample included 307 children and 298 parents from the original sample of participants. Data were collected on 300 children and 232 parents.

Children and Families

The children were 4.5 years old at the time of baseline data collection and slightly more than half (52%) were male. The sample of children was White (80%), African American (18%), and Hispanic (11%). Table 2.1 provides additional information on the demographic characteristics of the children in the study sample. At baseline, a higher percentage of control group parents reported that their child had an Individualized Education Plan (IEP) relative to those assigned to the *Bright Beginnings* and *Creative Curriculum* conditions (33% vs. 13% and 12%, $p < .01$).

Table 2.1. Child demographic characteristics for *Bright Beginnings* and *Creative Curriculum*

Characteristics	Full sample n = 309	Curriculum comparison		
		Control n = 105	Treatment 1 ¹ n = 101	Treatment 2 ² n = 103
Age at baseline (years), mean	4.5	4.5	4.6	4.5
Gender (% male)	52.1	48.6	53.5	54.4
Race/ethnicity (%)				
White, non-Hispanic	79.6	84.0	74.4	80.0
African American, non-Hispanic	6.5	‡	‡	‡
Hispanic	10.8	10.6	11.1	10.5
Asian or Pacific Islander	‡	‡	‡	‡
Native American	‡	‡	‡	‡
Multiple/other	‡	‡	‡	‡
Child disability status (parent reported, %)	18.7	32.5**	11.7	13.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

** $p < .01$

¹Treatment 1 is *Creative Curriculum*.

²Treatment 2 is *Bright Beginnings*.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 2.2. The average age of the primary caregiver was 31 years. More than half (65%) of the primary caregivers were married. Less than half reported having had some college (34%) or had graduated from college (7%), 38 percent had a high school diploma or GED, and 21 percent had not finished high school. Less than half (43%) of the primary caregivers were employed full-time, 13 percent were employed part-time, and 39 percent were unemployed. There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

Table 2.2. Primary caregiver demographic characteristics for *Bright Beginnings* and *Creative Curriculum*

Characteristics	Full sample n = 249	Curriculum comparison		
		Control n = 78	Treatment 1 ¹ n = 78	Treatment 2 ² n = 93
Age at baseline (years), mean	30.6	31.3	29.5	31.0
Marital status (%)				
Married	64.7	66.7	61.5	65.6
Separated/Divorced	18.5	17.9	17.9	19.4
Widowed	‡	‡	0.0	‡
Never Married	15.7	14.1	20.5	12.9
Race/ethnicity (%)				
White, non-Hispanic	83.5	93.6	76.9	80.4
African American, non-Hispanic	7.3	3.8	10.3	7.6
Hispanic	7.7	2.6	11.5	8.7
Asian or Pacific Islander	‡	0.0	0.0	‡
Native American	‡	0.0	0.0	‡
Multiple/other	‡	0.0	‡	‡
Educational level (%)				
Did not finish high school	21.4	12.8	25.6	25.0
High school diploma or GED	38.3	50.0	34.6	31.5
Some college	33.5	34.6	35.9	30.4
College graduate	6.9	‡	‡	13
Employment (%)				
Full-time	43.4	42.3	39.7	47.3
Part-time	12.9	16.7	10.3	11.8
Unemployed	39.0	37.2	44.9	35.5
Other	4.8	‡	5.1	5.4

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹Treatment 1 is *Creative Curriculum*.

²Treatment 2 is *Bright Beginnings*.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Teachers

There were 21 teachers who participated in the preschool year intervention study. All of the teachers were female, and all were White. On average, the preschool teachers had 11 years of teaching experience, with an average of 6 years of experience teaching preschool. All of the teachers had a bachelor's (52%) or graduate (48%) degree. All reported having a state-awarded teacher certification. Table 2.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Table 2.3. Preschool teacher characteristics for *Bright Beginnings* and *Creative Curriculum*

Characteristics	Full sample n = 21	Curriculum comparison		
		Control n = 7	Treatment 1 ¹ n = 7	Treatment 2 ² n = 7
Gender (% female)	100.0	100.0	100.0	100.0
Race/ethnicity (%)				
White, non-Hispanic	100.0	100.0	100.0	100.0
African American, non-Hispanic	0.0	0.0	0.0	0.0
Hispanic	0.0	0.0	0.0	0.0
Asian or Pacific Islander	0.0	0.0	0.0	0.0
Native American	0.0	0.0	0.0	0.0
Multiple/other	0.0	0.0	0.0	0.0
Educational level (%)				
High school diploma or GED	0.0	0.0	0.0	0.0
Associate's degree	0.0	0.0	0.0	0.0
Bachelor's degree	52.0	‡	‡	71.0
Graduate degree	48.0	57.0	57.0	‡
Current teaching license/certificate (%)	100.0	100.0	100.0	100.0
Child Development Associate (CDA) (%)	0.0	0.0	0.0	0.0
State-awarded preschool certificate (%)	0.0	0.0	0.0	0.0
No credential (%)	0.0	0.0	0.0	0.0
Years of teaching experience, overall (mean)	11.2	7.6	11.4	14.7
Years of preschool teaching experience (mean)	5.9	3.4	4.5	9.6

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹Treatment 1 is *Creative Curriculum*.

²Treatment 2 is *Bright Beginnings*.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Programs/Classrooms

The average preschool class size was 16.6 children. The child-staff ratio was on average 7.2 children to 1 teacher or program staff person.

Random Assignment

Randomization of 36 classrooms to the three curriculum conditions was done for the site-specific evaluation during the pilot year of curriculum implementation (2002-03). During the pilot year, 21 of those classrooms were randomly selected to also participate in the PCER initiative. For the second year of implementation (2003-04), 13 of those 21 classrooms continued to participate in the initiative. These included five

implementing *Bright Beginnings*, four implementing *Creative Curriculum*, and four control classrooms. The 8 classrooms that dropped out were replaced by other classrooms (randomly selected) from the original 36 classrooms. These included two implementing *Bright Beginnings*, three implementing *Creative Curriculum*, and three control classrooms. All the teachers in these 21 classrooms were the same as during the pilot year, except for one teacher in a control classroom, and she had substituted for the teacher in that classroom while the latter was on maternity leave.

For the initial randomization, a total of 36 state pre-kindergarten classrooms within 28 schools in 7 county school systems were assigned to treatment and control conditions at the beginning of the pilot year of the study. The preschool classrooms were blocked into groups of three by matching them as closely as possible on demographic and academic performance variables for the elementary school geographically nearest each preschool (in many cases the preschool was in the same school facility as the elementary school). The matching variables were derived from information available on the Tennessee State Department of Education website and consisted of two composite factors that were created for this purpose—a demographic factor (urban/rural and percentage of races other than White) and a composite achievement factor (percent free lunch and reading, language, mathematics, and science achievement test scores). Classrooms in the same school—no more than two in any instance—were included as a single unit in these blocks to ensure that they would not be assigned to different conditions. Within each block, one classroom (or pair, if two in the same school were a single unit in the block) was randomly assigned to the *Bright Beginnings* curriculum condition, one to the *Creative Curriculum* condition, and one to the control group condition, with the constraint that the classrooms in a given county school system be distributed over the three conditions. All 36 teachers and their assistants consented to participate in the pilot-year study. The evaluation for the PCER initiative was conducted on a subset of the larger site-specific sample of teachers, parents, and children. That subset consisted of a random selection of 21 of the 36 classrooms and 309 children in those classrooms.

Contamination

The 21 pre-kindergarten classrooms in the evaluation were divided across 18 schools. For 15 of those schools, there was only one classroom that participated in the study. For three schools, two classrooms participated, with both assigned to the same experimental condition. Consequently, the risk of contamination from teachers in different conditions exchanging curriculum information or materials was minimal.

Control Condition

In the control condition, teachers used teacher-developed, nonspecific curricula with a focus on basic school readiness.

Data Collection

RTI International (RTI) collected the child, teacher, and school data for the Tennessee site for all three waves of data collection. The Tennessee research team was responsible for conducting the parent interviews in the preschool year. In the kindergarten follow-up year, RTI staff completed the parent interviews.

The fall assessment data collection window for child assessments ranged from September 4, 2003 to November 7, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 8 days. The spring pre-kindergarten window was March 30, 2004 to May 11, 2004, and the kindergarten follow-up window was April 4, 2005 to June 24, 2005.

Attrition

Twenty-one classrooms were randomly assigned to treatment or control condition, all of which remained in the study throughout the pre-kindergarten year. For the child assessment, the baseline (fall 2003) response rate was 100 percent, the spring 2004 response rate was 95 percent, and the kindergarten follow-up response rate was 98 percent.

Implementation

The teacher sample included teachers who participated in the pilot year of the study (2002-03), and new teachers who started in 2003-04. The intervention teachers received 2.5 days of curriculum implementation training prior to the beginning of the school year. Teachers had access to ongoing curriculum implementation support throughout the school year. Onsite consultation to teachers was provided four times during the school year—twice by trained Tennessee research staff members, and twice by curriculum trainers. Consultation visits typically included a classroom observation, an opportunity for teachers to ask questions about the curriculum, and implementation feedback from the curriculum trainer and/or research assistant. Consultation visits were conducted in September, October, November to January, and late February.

The research team conducted site-specific curriculum fidelity classroom observations three times throughout the year in the fall (October to November 2003), winter (January to February 2004), and spring (April to May 2004). Visits were made to both treatment and control classrooms. Each classroom was visited for a full morning, followed by an interview with the teacher. The developers of each curriculum provided the researchers with a fidelity instrument used to assess implementation. Both instruments were used in all the classrooms, including the control classrooms. Both of the site-specific fidelity measures included items that addressed general early childhood practice, as well as items that focused on specific activities/practices/materials that were unique to each curriculum. For the *Bright Beginnings* intervention curriculum, all classrooms were rated on a four-point scale (weak, fair, good, excellent). The site-specific fidelity measures indicated that most classrooms were rated as showing a High or Medium level of implementation. One classroom was rated as Low on the fidelity measure. The *Creative Curriculum* classrooms were rated as High (2), Medium (2), Low (2), and Not at All (1) on the site-specific fidelity measure. The control classrooms received an average rating of Medium on the fidelity measure.

Implementation Fidelity Ratings

Bright Beginnings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. *Bright Beginnings* was rated Medium (1.88) on the global implementation fidelity measure. The control group curriculum was rated at the Medium (2.0) level as well.

Creative Curriculum

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. *Creative Curriculum* was rated Medium (2.14) on implementation fidelity. The control group curriculum was rated at the Medium (2.0) level as well.

Impact Analysis Results

We present analyses for each curriculum separately, beginning with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data.

Bright Beginnings—Child Outcomes

The unadjusted mean scores for the child-level measures are reported in table C-1a in appendix C. Covariate adjusted mean differences and standard errors are presented in table D-1a in appendix D. For all analyses of the measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 2.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically significant differences on these measures for the fall assessment.

There were no statistically detectable differences on the spring pre-kindergarten or kindergarten assessments on any of the mathematics assessments.

Based on the analyses for the three mathematics measures we conclude that *Bright Beginnings* did not have an effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically significant differences on these measures for the fall assessment.

There were no statistically detectable differences in the spring pre-kindergarten or kindergarten assessments on the WJ Letter Word Identification test or WJ Spelling test.

In spring of the pre-kindergarten year, there was a statistically reliable mean difference in scores on the TERA (ESs = .39, $p < .05$) favoring the *Bright Beginnings* group. No difference was found for the TERA for the spring kindergarten assessment.

Based on the analyses for the three reading measures, we conclude that *Bright Beginnings* did not have an effect on young children's early reading skills relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically significant difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were (a) the Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically significant difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Bright Beginnings* did not have an effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically significant differences on either measure for the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the two language measures, we conclude that *Bright Beginnings* did not have an effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically significant differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Bright Beginnings* did not have an effect on children's social and learning behaviors relative to the control condition.

Bright Beginnings—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-1b in appendix C. Covariate adjusted mean differences and standard errors are presented in table D-1b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 2.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was a statistically significant difference between groups for the fall observation ($ES_c = 1.39$, $p < .05$; follow-up analyses for this finding are included in appendix A). *Bright Beginnings* classrooms received higher global classroom quality ratings.

No statistically detectable differences between groups were obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Bright Beginnings* did not have an effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There was a

statistically significant difference at the time of the fall observation on the Arnett Detachment scale ($ES_c = -1.16, p < .05$; follow-up analyses for this finding are included in appendix A). Teachers in the *Bright Beginnings* classrooms were rated as being less detached in their interactions with their students relative to teachers in the control classrooms. There were no statistically significant differences on the other scales for the fall pre-kindergarten observation.

In spring of the kindergarten year, there were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interactions scales.

Based on the analyses of the four Arnett scales, we conclude that *Bright Beginnings* did not have an effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and TBRS Written Expression scales); (b) phonological awareness (TBRS Phonological Awareness scale); (c) language (TBRS Book Reading and Oral Language scales); and (d) early mathematics (TBRS Math Concepts scale) in spring of the pre-kindergarten year only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were statistically reliable effects favoring the *Bright Beginnings* classrooms on the Phonological Awareness ($ES_c = 1.53, p < .05$), Print and Letter Knowledge ($ES_c = 1.51, p < .05$), and Written Expression ($ES_c = 1.61, p < .01$) scales of the TBRS. There were no statistically detectable differences on the remaining scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Bright Beginnings* had a positive effect on early literacy and phonological awareness instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Bright Beginnings* had a positive effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Bright Beginnings* did not have a statistically detectable effect on language instruction.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Bright Beginnings* did not have a statistically detectable effect on early mathematics instruction.

Summary of Findings for *Bright Beginnings*

The findings for *Bright Beginnings* are summarized in table 2.4.

Table 2.4. Effect sizes for *Bright Beginnings*

Measure	Student-level effect sizes (ES _s)		
	RM analysis	RM analysis	ANCOVA
	Spring Pre-K	Spring K	Spring K
Mathematics			
WJ Applied Problems	.16	.13	—
CMA-A Mathematics Composite	.14	.07	—
Shape Composition ¹	-.03	.15	—
Reading			
TERA	.39*	-.07	—
WJ Letter Word Identification	.35	.09	—
WJ Spelling	.18	.06	—
Phonological awareness			
Pre-CTOPPP/CTOPP	-.07	†	.01
Language			
PPVT	.13	.07	—
TOLD	.09	.16	—
Behavior			
SSRS Social Skills	-.27	†	-.03
SSRS Problem Behavior ²	.23	†	.24
PLBS/LBS	.04	†	-.30
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis	ANCOVA	
	Spring Pre-K	Spring Pre-K	
Global classroom quality			
ECERS-R	.80	—	
Teacher-child interaction			
Arnett Detachment ³	.19	—	
Arnett Harshness ³	.12	—	
Arnett Permissiveness ³	.16	—	
Arnett Positive Interactions	.41	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	1.03	
TBRS Oral Language	†	.39	
TBRS Phonological Awareness	†	1.53*	
TBRS Print and Letter Knowledge	†	1.51*	
TBRS Written Expression	†	1.61*	
TBRS Math Concepts	†	.98	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$ ¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Creative Curriculum—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-2a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-2a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 2.5.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, CMA-A Composite Score, and Shape Composition). There were no statistically detectable differences at the fall assessment.

There were no statistically detectable differences on the spring pre-kindergarten or kindergarten assessments on any of the mathematics assessments.

Based on the analyses for the three mathematics measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (TERA, WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences at the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three reading measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on early reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Pre-CTOPPP, Elision subtest, and the CTOPP, Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Creative Curriculum* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences for the fall assessment.

There were no statistically detectable differences on either of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the two language measures, we conclude that *Creative Curriculum* did not have an effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Creative Curriculum* did not have an effect on children's social and learning behaviors relative to the control condition.

Creative Curriculum—Classroom Outcomes

The unadjusted mean scores for the classroom-level measures are reported in table C-2b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-2b in appendix D. For all analyses of the classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_C) are presented in table 2.5.

Overall classroom environment

We conducted a repeated measures analysis on the ECERS-R. There was a statistically significant difference between groups on the fall observation ($ES_C = 1.94, p < .01$); follow-up analyses for this finding are included in appendix A. *Creative Curriculum* teachers received higher overall classroom quality ratings.

No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Creative Curriculum* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There was a statistically significant difference at the time of the fall observation on the Arnett Detachment scale ($ES_C = -.95, p < .05$; follow-up analyses for this finding are included in appendix A). Teachers in the *Creative Curriculum* classrooms were rated as being less detached in their interactions with their students relative to teachers in the control classrooms. There were no statistically detectable differences on the other scales for the fall observation.

In spring of the pre-kindergarten year, there were no statistically significant differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales.

Based on the analyses of the four Arnett scales, we conclude that *Creative Curriculum* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (TBRS Print and Letter Knowledge and Written Expression scales); (b) phonological awareness (TBRS Phonological Awareness scale); (c) language (TBRS Book Reading and Oral Language scales); and (d) early mathematics (TBRS Math Concepts scale) in spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences between the treatment and control classrooms on the TBRS scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Creative Curriculum* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Creative Curriculum* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Creative Curriculum* did not have a statistically detectable effect on language instruction.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Creative Curriculum* did not have a statistically detectable effect on early mathematics instruction.

Summary of Findings for *Creative Curriculum*

The findings for *Creative Curriculum* are summarized in table 2.5.

Table 2.5. Effect sizes for *Creative Curriculum*: Tennessee

Measure	Student-level effect sizes (ES _s)		
	RM analysis	RM analysis	ANCOVA
	Spring Pre-K	Spring K	Spring K
Mathematics			
WJ Applied Problems	.17	.17	—
CMA-A Mathematics Composite	.10	.05	—
Shape Composition ¹	-.12	.00	—
Reading			
TERA	.02	.10	—
WJ Letter Word Identification	.16	.38	—
WJ Spelling	.19	.25	—
Phonological awareness			
Pre-CTOPPP/CTOPP	-.10	†	.06
Language			
PPVT	.23	.12	—
TOLD	.07	.11	—
Behavior			
SSRS Social Skills	-.03	†	.35
SSRS Problem Behavior ²	.07	†	-.05
PLBS/LBS	.14	†	.08
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis	ANCOVA	
	Spring Pre-K	Spring Pre-K	
Global classroom quality			
ECERS-R	.45	—	
Teacher-child interaction			
Arnett Detachment ³	-.16	—	
Arnett Harshness ³	-.12	—	
Arnett Permissiveness ³	.51	—	
Arnett Positive Interactions	-.15	—	
Teacher instructional practices ⁴			
TBRs Book Reading	†	-.47	
TBRs Oral Language	†	-.07	
TBRs Phonological Awareness	†	1.97	
TBRs Print and Letter Knowledge	†	1.81	
TBRs Written Expression	†	1.99	
TBRs Math Concepts	†	1.48	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRs measures did not include baseline pretest scores because TBRs data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (*p*-values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Chapter 3. *Creative Curriculum*: University of North Carolina at Charlotte (North Carolina and Georgia sites)

Curriculum

The research team at the University of North Carolina at Charlotte (North Carolina), implemented *Creative Curriculum*, a comprehensive curriculum for 3- to 5-year-old children.² The curriculum addresses four areas of development: social/emotional, physical, cognitive, and language development.

Creative Curriculum requires the physical space of the classroom to be structured into 10 interest areas: blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, and computers. Time is also allotted for outdoor activities. The 10 interest areas are designed to address curriculum content, such as literacy, mathematics, science, social studies, the arts, technology, and process skills, such as observing, exploring, and problem solving.

Sample

The North Carolina research team recruited Head Start programs in North Carolina and Georgia. All of the programs were full-day programs. Head Start teachers, assistants, and site managers were offered a stipend for participating in the study. Eight classrooms in North Carolina and 10 classrooms in Georgia were recruited to participate in the study. The parental consent process began before the start of the school year. The North Carolina research team relied on teachers to assist them in recruiting parents for the study. Teachers were given a letter to give to the parents during their initial home visit with the parents. Any parents who did not participate in a home visit were given a letter when he or she first came to the school. A sample of 18 classrooms and 194 children (97 treatment, 97 control) and parents were recruited for participation in the study. Data were collected on 190 children and 168 parents at the time of the fall baseline data collection.

In the follow-up year of the evaluation, the sample of schools went from five in pre-kindergarten to more than 54 schools in kindergarten. The sample of classrooms went from 18 preschool to 122 kindergarten classrooms. The kindergarten sample included 190 children and parents from the original sample of participants. Data were collected on 162 children and 135 parents.

Children and Families

The children were 4.5 years old at the time of baseline data collection and less than half were male (46%). The majority of the children were African American (85%). Table 3.1 provides additional information on the demographic characteristics of the children in the study sample. There were no statistically detectable differences between the treatment and control groups on these child characteristics.

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 3.2. The average age of the primary caregiver was 32 years. Less than half (39%) were married, and 39 percent were never married. Almost half of the primary caregivers reported having had some college (42%) or had graduated from college (6%), 29 percent had a high school diploma or GED, and 23 percent had not finished high school. Less than half (45%) of the primary caregivers were employed full-time, 10 percent were employed part-time, and 41 percent were unemployed. There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

² The Vanderbilt University (Tennessee) research team also evaluated *Creative Curriculum*.

Table 3.1. Child demographic characteristics for *Creative Curriculum*: North Carolina and Georgia

Characteristics	Full sample n = 194	Curriculum comparison	
		Control n = 97	Treatment n = 97
<i>Creative Curriculum: North Carolina and Georgia</i>			
Age at baseline (years), mean	4.5	4.5	4.5
Gender (% male)	45.8	47.4	44.2
Race/ethnicity (%)			
White, non-Hispanic	2.9	‡	‡
African American, non-Hispanic	85.0	84.1	85.9
Hispanic	7.5	6.8	8.2
Asian or Pacific Islander	‡	0.0	
Native American	0.0	0.0	0.0
Multiple/other	4.0	5.7	‡
Child disability status (parent reported, %)	20.1	23.9	16.0
Characteristics	Full sample n = 97	Curriculum comparison	
		Control n = 48	Treatment n = 49
<i>Creative Curriculum: North Carolina</i>			
Age at baseline (years), mean	4.5	4.5	4.5
Gender (% male)	46.4	47.9	44.9
Race/ethnicity (%)			
White, non-Hispanic	5.5	‡	‡
African American, non-Hispanic	81.3	80.9	81.8
Hispanic	11.0	10.6	11.4
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	‡	‡	0.0
Child disability status (parent reported, %)	28.7	30.4	26.8
Characteristics	Full sample n = 93	Curriculum comparison	
		Control n = 47	Treatment n = 46
<i>Creative Curriculum: Georgia</i>			
Age at baseline (years), mean	4.5	4.5	4.5
Gender (% male)	45.2	46.8	43.5
Race/ethnicity (%)			
White, non-Hispanic	0.0	0.0	0.0
African American, non-Hispanic	89.0	87.8	90.2
Hispanic	‡	‡	‡
Asian or Pacific Islander	‡	0.0	‡
Native American	0.0	0.0	0.0
Multiple/other	6.1	9.8	‡
Child disability status (parent reported, %)	11.0	16.7	5.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 3.2. Primary caregiver demographic characteristics for *Creative Curriculum*: North Carolina and Georgia

Characteristics	Full sample n = 169	Curriculum comparison	
		Control n = 88	Treatment n = 81
Age at baseline (years), mean	31.8	32.4	31.1
Marital status (%)			
Married	38.5	42.0	34.6
Separated/Divorced	21.9	19.3	24.7
Widowed	‡	‡	0.0
Never Married	39.1	37.5	40.7
Race/ethnicity (%)			
White, non-Hispanic	4.1	‡	4.9
African American, non-Hispanic	81.7	83.0	80.2
Hispanic	7.1	5.7	8.6
Asian or Pacific Islander	‡	0.0	‡
Native American	0.0	0.0	0.0
Multiple/other	6.5	8.0	4.9
Educational level (%)			
Did not finish high school	22.8	20.9	24.7
High school diploma or GED	29.3	30.2	28.4
Some college	41.9	39.5	44.4
College graduate	6.0	9.3	‡
Employment (%)			
Full-time	45.0	42	48.1
Part-time	10.1	10.2	9.9
Unemployed	40.8	40.9	40.7
Other	4.1	6.8	‡

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Teachers

There were 18 teachers who participated in the preschool year intervention study. All of the teachers were female. The majority of the teachers were African American (89%). The preschool teachers had on average 12 years of teaching experience, with an average of 9 years of experience teaching preschool. Half (50%) had an associate's degree, and 44 percent had a high school diploma or GED. Seventy-two percent of the teachers had a state-awarded teacher certification, and 78 percent had a Child Development Associate (CDA) credential. Table 3.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 14.4 children. The child-staff ratio was on average 7.6 children to one teacher or program staff person.

Table 3.3. Preschool teacher characteristics for *Creative Curriculum*: North Carolina and Georgia

Characteristics	Full sample n = 18	Curriculum comparison	
		Control n = 9	Treatment n = 9
Gender (% female)	100.0	100.0	100.0
Race/ethnicity (%)			
White, non-Hispanic	‡	‡	0.0
African American, non-Hispanic	89.0	78.0	100.0
Hispanic	0.0	0.0	0.0
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	6.0	11.0	0.0
Educational level (%)			
High school diploma or GED	44.0	67.0	‡
Associate's degree	50.0	‡	67.0
Bachelor's degree	‡	0.0	11.0
Graduate degree	0.0	0.0	0.0
Current teaching license/certificate (%)	24.0	0.0	44.0
Child Development Associate (CDA) (%)	78.0	78.0	78.0
State-awarded preschool certificate (%)	72.0	67.0	8.0
No credential (%)	0.0	0.0	0.0
Years of teaching experience, overall (mean)	11.9	12.9	11.0
Years of preschool teaching experience (mean)	8.7	9.2	8.2

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Random Assignment

Randomization of teachers within centers was done during the pilot year (2002-03) of the study, and the same assignments were maintained for the second year (2003-04) of curriculum implementation. At the end of the pilot year, the North Carolina site retained eight (four treatment and four control) of the 10 classrooms. Two classrooms were dropped because they were funded by the state's More at Four program, had degreed teachers, and had problems with high rates of teacher attrition. The Georgia site retained 10 out of 10 classrooms.

Treatment and control classrooms were housed in the same centers. Teachers were randomly assigned to either the treatment or control condition. The teachers within each center were assigned to blocks based on educational level and teacher certification status. They were then randomly assigned to treatment and control conditions within blocks. Children were randomly assigned to classrooms within each center. The children were randomly assigned to classrooms by blocking within each center by gender, special needs status, and ethnicity. Children were then randomly assigned to classrooms within blocks.

Each of these randomization procedures was conducted by using a pseudorandom number-generating software program to assign a random number to each participant. A random seed was set each time a new batch of numbers was generated. The participants were then sorted by their random number to make assignments to conditions. This process ensured that each participant (teacher or child) within each block had the same probability of assignment to treatment or control conditions.

A total of five Head Start preschools in two states (North Carolina and Georgia) were purposefully selected in the fall of the preschool year by the research team. There were three participating centers in North Carolina and two in Georgia. Randomization to either *Creative Curriculum* or the control curriculum was carried out at the classroom level and at child levels as described above. A total of 18 classrooms and 194 children took part in the study.

Contamination

Treatment and control classrooms were housed in the same centers. Teachers within the Head Start centers worked closely together. There may have been a few instances where a treatment group teacher inadvertently shared aspects from the treatment curriculum content or training with a control teacher. The research team conducted focus groups of both groups of teachers and there were not many comments about sharing.

The site-specific implementation fidelity data suggests that the control group teachers were doing some of the activities on the *Creative Curriculum* fidelity checklist. However, many of the items focused on generally accepted early childhood practice and were not curriculum-specific. It is not possible to attribute the control group scores to a contamination effect.

Control Condition

In the control condition, teachers used teacher-developed, nonspecific curricula.

Data Collection

RTI International (RTI) collected the child, teacher, and school data for the North Carolina and Georgia sites for all three waves of data collection. The North Carolina research team was responsible for conducting the parent interviews in the preschool year. In the kindergarten follow-up year, RTI staff completed the parent interviews. The fall assessment data collection window for child assessments ranged from September 3, 2003 to October 24, 2003 in North Carolina and August 25, 2003 to October 15, 2003 in Georgia. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 16 days in North Carolina and 14 days in Georgia. The spring pre-kindergarten window was March 3, 2004 to May 11, 2004 in North Carolina and April 13, 2004 to June 19, 2004 in Georgia. The kindergarten follow-up window was April 4, 2005 to May 24, 2005 in North Carolina and March 28, 2005 to June 30, 2005 in Georgia.

Attrition

Eighteen classrooms were randomly assigned to treatment and control conditions. All 18 classrooms remained in the study throughout the pre-kindergarten year.

For the child assessment, the baseline (fall 2003) response rate was 98 percent, the spring 2004 response rate was 90 percent, and the kindergarten follow-up response rate was 85 percent.

Implementation

Teachers in the treatment group were in their second year of implementing *Creative Curriculum* at the time of the evaluation. The North Carolina research team provided refresher training to the treatment group teachers during the second year of curriculum implementation. The training was delivered in 1-day or half-day sessions that were offered between August 2003 and February 2004. Four training sessions were provided to the treatment group teachers in North Carolina. Five training sessions were provided to the treatment group teachers in Georgia. Training topics included choosing and planning in-depth topics of study with children; providing materials and interactions for content learning (literacy, mathematics, science, social studies, the arts, and technology); and observation-based assessment of children's learning. The training was designed as a mix of lecture, small group projects, video viewing, and hands-on practical application activities. The same

trainer conducted training in both North Carolina and Georgia. Technical assistance was provided to teachers on an ongoing basis throughout the school year. The *Creative Curriculum* trainer and the local site coordinators at each project location provided technical assistance. Technical assistance was provided from August/September 2003 to April 2004.

An independent observer collected implementation fidelity data in fall of the preschool year. Observations were conducted in both treatment and control classrooms. The fidelity measure domains included the physical environment of the classrooms, the structure of the classrooms, teacher-child interactions, assessment, and family involvement. On average, the treatment group scored 86 percent on the implementation checklist and the control group scored approximately 58 percent on the implementation checklist. A *Creative Curriculum* classroom is considered as meeting the publishers' implementation criteria if the scores are greater than or equal to 80 percent at the scale level (e.g., teacher-child interactions or assessment) and 85 percent on the total score. At the time of the fall 2003 classroom fidelity observations, only two of nine treatment classrooms reached this level of implementation. None of the control classrooms met the criteria. The spring 2004 fidelity observations indicated that six of nine treatment classrooms had reached the implementation criteria. None of the control classrooms reached the 85 percent total score criterion.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. *Creative Curriculum* was rated Medium (2.11) on the global implementation fidelity measure. The control group curriculum was rated at the low Medium (1.5) level.

Impact Analysis Results

We present analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, language, and behavioral assessments) followed by the analyses of the classroom observation data. Our discussion of the results focuses on the combined analyses of the Georgia and North Carolina sites.

Creative Curriculum—Child Outcomes

The unadjusted mean scores for the child-level measures are reported in table C-3a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-3a in appendix D. For all analyses of the child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 3.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences for the fall assessment.

There were no statistically detectable differences on the spring pre-kindergarten or kindergarten assessments on any of the mathematics assessments.

Based on the analyses of the three mathematics measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences for the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the three reading measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were (a) the Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Creative Curriculum* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences for the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the two language measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Creative Curriculum—Classroom Outcomes

The unadjusted mean scores for classroom-level measures are reported in table C-3b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-3b in appendix D. For all analyses of classroom-level measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_C) are presented table 3.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups for the fall observation.

There was a statistically significant difference between the *Creative Curriculum* classrooms and the control classrooms on the ECERS-R for the spring pre-kindergarten observation ($ES_C = 1.66, p < .05$). Treatment group classrooms received higher overall classroom quality ratings relative to the control group classrooms.

Based on the analysis of the ECERS-R, we conclude that *Creative Curriculum* had a positive effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

In spring of the pre-kindergarten year, there were statistically significant differences on the Arnett Detachment ($ES_C = -1.68, p < .05$) scale, indicating that teachers in the *Creative Curriculum* classrooms were less detached in their interactions with their students relative to teachers in the control classrooms. There was also a statistically significant effect on the Positive Interactions scale ($ES_C = 1.65, p < .01$), indicating that teachers in the *Creative Curriculum* classrooms were observed having more positive interactions with children relative to teachers in the control classrooms. No statistically detectable differences were obtained on the Arnett Harshness scale.

Based on the analyses of the four Arnett scales, we conclude that *Creative Curriculum* had a positive effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) in the spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted; the covariates were (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were statistically reliable differences favoring the *Creative Curriculum* classrooms on the Written Expression ($ES_C = 1.73, p < .01$) and Oral Language ($ES_C = 1.80, p < .01$) scales. There were no statistically significant differences on the Book Reading, Print and Letter Knowledge, Phonological Awareness, or Math Concepts scales of the TBRS.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Creative Curriculum* had a positive effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Creative Curriculum* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Creative Curriculum* had a positive effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Creative Curriculum* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Creative Curriculum*: North Carolina and Georgia

The findings for *Creative Curriculum* are summarized in table 3.4.

Table 3.4. Effect sizes for *Creative Curriculum*: North Carolina and Georgia

Measure	Student-level effect sizes (ES_s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.20	.09	—
CMA-A Mathematics Composite	-.10	.14	—
Shape Composition ¹	.19	-.01	—
Reading			
TERA	-.08	-.04	—
WJ Letter Word Identification	-.08	.00	—
WJ Spelling	-.18	-.05	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.02	†	.06
Language			
PPVT	.08	.15	—
TOLD	-.16	-.17	—
Behavior			
SSRS Social Skills	.05	†	-.12
SSRS Problem Behavior ²	-.16	†	.08
PLBS/LBS	.07	†	-.20
Measure	Classroom-level effect sizes (ES_c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	1.66*	—	
Teacher-child interaction			
Arnett Detachment ³	-1.68*	—	
Arnett Harshness ³	-.70	—	
Arnett Permissiveness ³	-1.01	—	
Arnett Positive Interactions	1.65**	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	.28	
TBRS Oral Language	†	1.80**	
TBRS Phonological Awareness	†	-.10	
TBRS Print and Letter Knowledge	†	1.02	
TBRS Written Expression	†	1.73**	
TBRS Math Concepts	†	.75	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$ ¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Chapter 4. Creative Curriculum with Ladders to Literacy: University of New Hampshire (New Hampshire site)

Curriculum

The University of New Hampshire (New Hampshire) research team chose to evaluate *Ladders to Literacy*, an early literacy and language development supplementary curriculum for preschool and kindergarten children. The curriculum is intended for use in inclusive classrooms with children with disabilities and children for whom English is a second language. The curriculum includes more than 50 skill-building activities that are organized into three sections:

- print awareness;
- metalinguistic awareness; and
- oral language.

The activities included in the curriculum are designed to be suggestions or models that teachers can adopt for use with an existing classroom curriculum. Teachers are encouraged to select the activities that they want to implement and incorporate those activities into their daily classroom schedule. Teachers are provided with guidance on how to use scaffolding techniques to individualize children's learning of language and literacy skills.

The New Hampshire researchers selected a common subset of 27 activities that all *Ladders to Literacy* treatment group teachers used throughout the school year. For this evaluation, *Ladders to Literacy* was implemented as a supplementary curriculum to the *Creative Curriculum*. Classrooms in the control condition implemented the *Creative Curriculum* without the supplement.

Sample

The New Hampshire research team recruited Head Start classrooms to participate in the study. The Head Start program administrators and teachers received a program incentive for participating in the study. Less than half were full-day programs. A total of 14 teachers/classrooms were recruited for the study. The New Hampshire research team relied on the teachers in each of the participating classrooms to distribute consent forms to the families of eligible children. A sample of 123 children (62 treatment, 61 control) and parents were recruited to participate in the study. Data were collected on a total of 123 children and 20 parents at the time of the fall baseline data collection.

In the follow-up year of the evaluation study, the sample of schools went from 8 in pre-kindergarten to 26 schools in kindergarten. The sample of classrooms went from 14 preschool to 41 classrooms in kindergarten.

Children and Families

The children were 4.6 years old at the time of baseline data collection and less than half (44%) were male. The racial/ethnic composition of the sample of children was diverse: 39 percent White, 11 percent African American, and 31 percent identified as Hispanic. Table 4.1 provides additional information on the demographic characteristics of the children in the study sample. There were no statistically detectable differences between the treatment and control groups on these child characteristics.

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 4.2. The average age of the primary caregiver was 30 years. Less than half

Table 4.1. Child demographic characteristics for *Creative Curriculum with Ladders to Literacy*

Characteristics	Full sample n = 123	Curriculum comparison	
		Control n = 61	Treatment n = 62
Age at baseline (years), mean	4.6	4.6	4.6
Gender (% male)	43.9	41.0	46.8
Race/ethnicity (%)			
White, non-Hispanic	38.7	44.8	33.3
African American, non-Hispanic	11.3	‡	12.1
Hispanic	30.6	20.7	39.4
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	19.4	24.1	15.2
Child disability status (parent reported, %) ¹	25.0	0.0	33.3

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹ Because there were so few parent interviews in New Hampshire during the baseline round, the value reported in this table might be invalid. The parent interview rates were higher at the time of the spring pre-kindergarten interview (55 parent interviews) and the spring kindergarten interview (63 parent interviews).

NOTE: Child disability status was included in the child-level impact analyses. The analysis results should be interpreted with caution because of the questionable validity of the parent report data.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 4.2. Primary caregiver demographic characteristics for *Creative Curriculum with Ladders to Literacy*

Characteristics	Full sample n = 20 ¹	Curriculum comparison	
		Control n = 5	Treatment n = 15
Age at baseline (years), mean	30.2	31.8	29.7
Marital status (%)			
Married	40.0	40.0	40.0
Separated/Divorced	30.0	40.0	26.7
Widowed	0.0	0.0	0.0
Never Married	30.0	20.0	33.3
Race/ethnicity (%)			
White, non-Hispanic	50.0	80.0	40.0
African American, non-Hispanic	0.0	0.0	0.0
Hispanic	20.0	0.0	26.7
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	30.0	‡	33.3
Educational level (%)			
Did not finish high school	45.0	‡	46.7
High school diploma or GED	40.0	‡	33.3
Some college	‡	0.0	‡
College graduate	‡	0.0	‡
Employment (%)			
Full-time	30.0	‡	33.3
Part-time	0.0	0.0	0.0
Unemployed	70.0	80.0	66.7
Other	0.0	0.0	0.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹ Because there were so few parent interviews in New Hampshire during the baseline round, the values reported in the tables might be invalid. The parent interview rates were higher at the time of the spring pre-kindergarten interview (55 parent interviews) and the spring kindergarten interview (63 parent interviews). Mother's educational level was included in the child-level impact analyses. The analysis results should be interpreted with caution because of the questionable validity of the parent report data.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

(40%) were married, 30 percent were never married, and 30 percent were separated or divorced. Few (15%) primary caregivers reported having had some college or a bachelor's degree, 40 percent had a high school diploma or GED, and 45 percent had not finished high school. About one-third (30%) of the primary caregivers were employed full-time, and 70 percent were not employed. There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

Teachers

Fourteen teachers participated in the preschool year intervention study. Most of the preschool teachers were female. Most (93%) of the teachers were White. The preschool teachers had an average of 9 years of teaching experience and, on average, 8 years of experience teaching preschool. Nearly half of the teachers had either a bachelor's degree (36%) or higher, 29 percent had an associate's degree, and 29 percent had a high school diploma or GED. Forty-three percent of the teachers reported having a current teaching license or certification and 36 percent had no teacher certification credentials. Table 4.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 13.3 children. The child-staff ratio was on average 5.8 children to one teacher or program staff person.

Random Assignment

Randomization was done during the pilot year (2002-03) of curriculum implementation, and the assignments (with some modifications) were maintained for the second year of implementation (2003-04). The research team used a random number software program to assign classrooms to conditions. In 2002, the New Hampshire research team selected 12 classrooms/teachers from a list of prospective study participants. They randomly selected four urban full-day classrooms from the list and randomly assigned two of those classrooms to the treatment condition and two to the control condition. The research team purposefully selected two urban half-day classrooms as a matched pair, with similarly high numbers of Spanish-speaking children who were enrolled in these classrooms. These two half-day classrooms were randomly assigned to the treatment and control conditions (one classroom in each condition). The research team randomly selected two additional urban half-day classrooms from the remaining list, and randomly assigned one of these classrooms to the control condition and the other to the treatment condition. The team grouped four suburban/rural classrooms that were on the list by the existence or absence of a kindergarten program in their respective towns. The research team randomly assigned one of the "kindergarten" towns and one of the "no kindergarten" towns to the control condition, and the other two classrooms to the treatment condition. A total of four full-day urban classrooms (two treatment and two control); four half-day urban classrooms (two control and two treatment); and four part-day rural classrooms (two control and two treatment) were included in the pilot-year study sample.

In the second year of implementation, the research team was able to retain 11 of the pilot-year classrooms. Two additional classrooms were added to the sample in the second year. The researchers flipped a coin to assign these two classrooms to treatment and control conditions. All of the six teachers in the treatment classrooms in the pilot year remained in the sample in the second year. Teachers were retained in three of the six control classrooms from the pilot year. One of the six control group teachers declined to participate in year 2 and was replaced by another classroom in the same center. In another two control group classrooms, the pilot-year teachers left the programs and the replacement teachers in these two classrooms agreed to participate in the study. A total of 14 classrooms and 123 children took part in the study. Details regarding randomization procedures and changes from the pilot year to the second year are described in chapter 1 of this report.

Table 4.3. Preschool teacher characteristics for *Creative Curriculum with Ladders to Literacy*

Characteristics	Full sample n = 14	Curriculum comparison	
		Control n = 7	Treatment n = 7
Gender (% female)	93.0	100.0	86.0
Race/ethnicity (%)			
White, non-Hispanic	93.0	86.0	100.0
African American, non-Hispanic	‡	‡	0.0
Hispanic	0.0	0.0	0.0
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	0.0	0.0	0.0
Educational level (%)			
High school diploma or GED	29.0	‡	‡
Associate's degree	29.0	‡	‡
Bachelor's degree	36.0	‡	57.0
Graduate degree	7.0	‡	0.0
Current teaching license/certificate (%)	43.0	71.0	‡
Child Development Associate (CDA) (%)	‡	0.0	‡
State-awarded preschool certificate (%)	‡	0.0	‡
No credential (%)	36.0	‡	57.0
Years of teaching experience, overall (mean)	9.3	7.6	11.0
Years of preschool teaching experience (mean)	7.6	6.7	8.4

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Contamination

In one of the sites, there were multiple classrooms (four classrooms) of both conditions in the same building. The research team met with the treatment teachers and aides to ensure that they were not sharing any materials or activities with the control group teachers. The research team also conducted classroom observations in the treatment and control classrooms during the preschool year. There was no evidence of contamination.

Control Condition

In the control classrooms, *Creative Curriculum* was implemented as it is normally implemented (i.e., without the *Ladders to Literacy* add-on). All of the teachers received at least 1 day of *Creative Curriculum* training from a staff member at Teaching Strategies, Inc.

Data Collection

RTI International (RTI) collected the child, teacher, and school data for the New Hampshire site for all three waves of data collection. The New Hampshire research team was responsible for conducting the parent interviews in the preschool year. In the kindergarten follow-up year, RTI staff completed the parent interviews. The fall assessment data collection window for child assessments ranged from October 6, 2003 to November 24, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 10 days. The spring pre-kindergarten window was April 2, 2004 to June 10, 2004 and the kindergarten follow-up window was April 21, 2005 to June 24, 2005.

Attrition

Fourteen classrooms were randomly assigned to treatment or control condition. All 14 classrooms remained in the study throughout the pre-kindergarten year.

For the child assessment, the baseline (fall 2003) response rate was 100 percent; the spring 2004 response rate was 85 percent; and the kindergarten follow-up response rate was 80 percent.

Implementation

The New Hampshire research team implemented the *Ladders to Literacy* supplementary curriculum as an add-on to the existing curriculum (*Creative Curriculum*). Thus, *Creative Curriculum* was in use in both treatment and control classrooms. All of the teachers received at least 1 day of *Creative Curriculum* training from a staff member at Teaching Strategies, Inc.

The treatment group teachers received initial *Ladders to Literacy* training in September of the preschool year. The treatment group teachers received ongoing *Ladders to Literacy* training on a monthly basis throughout the school year (October 2003 to April 2004). The treatment group teachers received training to implement 27 language and literacy activities that covered three domains (print/book awareness, metalinguistic awareness, and oral language). Teachers were expected to implement nine activities (three from each of the three major domains) in the months of November and December 2003. Teachers were instructed to cumulatively add three to six additional activities on a monthly basis, from January to May 2004, following an implementation schedule that went through May of the preschool year.

Site-specific curriculum fidelity observations were conducted in all of the 14 participating classrooms on a monthly basis from December through April of the preschool year. A total of 70 observations were made during that time period. In the treatment classrooms, observations were conducted during *Ladders to Literacy* activities. In the control classrooms, observations were conducted during activities similar in form to the *Ladders to Literacy* activities (e.g., morning circle time and story reading time). A second observer conducted an observation at the same time as a primary observer on 12 occasions. The New Hampshire research team evaluated inter-observer reliability based on 17 percent of the total number of observations conducted across the preschool year. The research team observed implementation of 23 of the 27 *Ladders to Literacy* activities at least once during the year. The site-specific fidelity observations indicated that teachers in the treatment classrooms implemented more items across *Ladders to Literacy* activities than teachers in the control classrooms.

In addition to conducting *Ladders to Literacy* implementation fidelity observations, the New Hampshire research team also conducted *Creative Curriculum* fidelity observations. The research team conducted *Creative Curriculum* fidelity observations in both the treatment and control classrooms. The *Creative Curriculum* implementation fidelity checklist included items in five areas: (1) physical environment of each activity area; (2) structure (e.g., daily schedules and routines); (3) teacher-child interactions; (4) assessment (e.g., children's progress); and (5) family involvement. The New Hampshire research team collected data on physical environment, structure, and teacher-child interactions. The research team conducted two *Creative Curriculum* fidelity observations across the 14 classrooms. The mean total proportion of implementation observed was 60 percent.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from "Not at All" (0) to "High" (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group

curriculum. The *Ladders to Literacy* curriculum was rated in the high Medium range (2.71) on the global implementation fidelity measure. The control group curriculum was rated at the Medium level (2.0).

Impact Analysis Results

We present analyses for each curriculum separately, beginning with the analyses of the child-level measures (mathematics, reading, phonological awareness, and language assessments), and followed by the analyses of the classroom observation data.

Creative Curriculum with Ladders to Literacy—Child Outcomes

The unadjusted mean scores for the child-level measures are reported in table C-4a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-4a in appendix D. For all analyses of child-level measures, the following covariates were included (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 4.4. There was a very low parent interview rate (20 interviews from a sample of 123) in New Hampshire during the baseline round of data collection. The parent interview rates were higher at the time of the spring pre-kindergarten interview (55 parent interviews) and the spring kindergarten interview (63 parent interviews). Child and parent background characteristics were included in the child level impact analysis models as covariates. The analysis results should be interpreted with caution because of low parent interview response rate and the questionable validity of the parent report data.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from the three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences on the spring pre-kindergarten or kindergarten assessments on any of the mathematics assessments.

Based on the analyses of the three mathematics measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences on these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the three reading measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically significant difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the Pre-CTOPPP fall assessment score, child's gender, age, race/ethnicity, disability status as reported by parent, and mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences on these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the two language measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behavior Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures at the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Creative Curriculum with Ladders to Literacy—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-4b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-4b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (E_{Sc}) are presented in table 4.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups on the fall observation.

No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

There were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales.

Based on the analyses of the four Arnett scales, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) in spring of the pre-kindergarten year only. To analyze these data, ANCOVAs were conducted; the covariates were (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There was a statistically reliable difference favoring the *Creative Curriculum with Ladders to Literacy* classrooms on Written Expression ($ESc = 1.13, p < .05$). There were no statistically detectable differences on the TBRS Book Reading, Print and Letter Knowledge, Oral Language, Phonological Awareness, or Math Concepts scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Ladders to Literacy* had a positive effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Creative Curriculum with Ladders to Literacy*

The findings for *Creative Curriculum with Ladders to Literacy* are summarized in table 4.4.

Table 4.4. Effect sizes for *Creative Curriculum with Ladders to Literacy*

Measure	Student-level effect sizes (ES_s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	-.14	-.33	—
CMA-A Mathematics Composite	.18	-.19	—
Shape Composition ¹	.02	-.10	—
Reading			
TERA	-.30	-.54	—
WJ Letter Word Identification	-.16	-.27	—
WJ Spelling	.30	-.08	—
Phonological awareness			
Pre-CTOPPP/CTOPP	-.16	†	-.10
Language			
PPVT	-.38	-.30	—
TOLD	-.22	-.06	—
Behavior			
SSRS Social Skills	-.25	†	.17
SSRS Problem Behavior ²	-.01	†	.02
PLBS/LBS	-.08	†	-.11
	Classroom-level effect sizes (ES_c)		
Measure	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	-.71	—	
Teacher-child interaction			
Arnett Detachment ³	.51	—	
Arnett Harshness ³	-.26	—	
Arnett Permissiveness ³	1.02	—	
Arnett Positive Interactions	.03	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	-.32	
TBRS Oral Language	†	-.50	
TBRS Phonological Awareness	†	-.19	
TBRS Print and Letter Knowledge	†	.75	
TBRS Written Expression	†	1.13*	
TBRS Math Concepts	†	.44	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$ ¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. There was a very low parent interview rate (20 interviews from a sample of 123) in New Hampshire during the baseline round of data collection. The parent interview rates were higher at the time of the spring pre-kindergarten interview (55 parent interviews) and the spring kindergarten interview (63 parent interviews). Child and parent background characteristics were included in the child-level impact analysis models as covariates. The analysis results should be interpreted with caution because of the low parent interview response rate and the questionable validity of the parent report data. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Page intentionally left blank.

Chapter 5. *Curiosity Corner*: Success for All Foundation (SFA sites: Florida, Kansas, and New Jersey)

Curriculum

The Success for All Foundation (SFA) research team evaluated *Curiosity Corner*, a comprehensive curriculum for 3- and 4-year-old children that was developed by SFA. The curriculum uses 38 thematic units to cover topics such as family life, opposites, seasons, and nature.

Each thematic unit includes suggested activities that are designed to promote children's language and literacy, and cognitive, mathematical, social, personal, creative, and physical development. Daily learning activities are built around learning labs, where children explore the theme through hands-on experiences and interaction with teachers. Teachers receive initial training and ongoing professional development support throughout the school year. The curriculum also features a home component, which provides families with a lending library, videos, and the opportunity for participation in classroom activities.

Sample

The SFA research team recruited preschool programs in three different states (Florida, Kansas, and New Jersey). Schools were recruited through phone calls from SFA researchers. The SFA research team targeted districts with SFA schools with preschool classes to fit their two (preschool curriculum types) x two (SFA and non-SFA kindergarten classrooms) study design. Children in the *Curiosity Corner* and control conditions in the preschool year transitioned into SFA and non-SFA schools during the kindergarten year of the study. The researchers first recruited SFA schools within a district then they asked for recommendations of non-SFA schools to participate in the study. When non-SFA schools with preschool programs were not available, the research team recruited Head Start and day care centers. The final sample included 31 teachers and classrooms. Parents were recruited with assistance from the preschool teachers, and offered an incentive to participate in the study. The average parental consent rate was 63 percent for the SFA-Florida site (61% for the treatment group, 66% for the control group); 77 percent for the SFA-Kansas site (70% for the treatment group, 83% for the control group); and 47 percent for the SFA-New Jersey site (59% for the treatment group, 38% for the control group). Across all three locations, 215 children (105 treatment, 110 control) and parents were recruited. Data were collected on a total of 211 children and 195 parents at the time of the fall assessment.

In the follow-up year of the evaluation, the sample of schools went from 18 in pre-kindergarten to 69 schools in kindergarten. The sample of classrooms went from 31 preschool to 107 kindergarten classrooms. Data were collected on 194 children and 184 parents from the original participant sample.

Children and Families

The children were 4.7 years of age at the time of baseline data collection and half (49%) were male. The sample primarily included African American (51%) and White (28%) preschoolers. The racial/ethnic composition of the sample of children varied based on the geographic location of the sample. Table 5.1 provides additional information on the demographic characteristics of the study sample. At baseline, a higher percentage of boys were in the *Curiosity Corner* classrooms relative to those assigned to the control group (61% vs. 38%, $p < .001$).

Table 5.1. Child demographic characteristics for *Curiosity Corner*

Characteristics	Full sample n = 215	Curriculum comparison	
		Control n = 110	Treatment n = 105
Age at baseline (years), mean	4.7	4.7	4.6
Gender (% male)	49.5	37.9	61.0***
Race/ethnicity (%)			
White, non-Hispanic	27.5	30.7	24.3
African American, non-Hispanic	50.5	42.6	58.3
Hispanic	13.7	20.8	6.8
Asian or Pacific Islander	2.9	0.0	5.8
Native American	‡	‡	0.0
Multiple/other	4.9	5.0	4.9
Child disability status (parent reported, %)	14.4	16.5	12.4

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

*** $p < .001$

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 5.2. The average age of the primary caregiver was 31 years. Approximately one-third (35%) of the primary caregivers were not married, and 46 percent were married at the time of the baseline data collection. Half (50%) of the primary caregivers reported having had some college or a college degree, 32 percent had a high school diploma or GED, and 18 percent had not finished high school. Approximately half (51%) of the primary caregivers were employed full-time, 13 percent were employed part-time, and 36 percent were not working at the time of the fall data collection. There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

Teachers

There were 31 teachers who participated in the preschool year intervention study. Almost all (97%) of the preschool teachers were female. The racial/ethnic composition of the sample included White (58%) and African American (19%) teachers. The preschool teachers had, on average, 10 years of teaching experience, and an average of 6 years of experience teaching preschool. The majority of teachers had a bachelor's (45%) or graduate (32%) degree. Sixteen percent of the teachers had an associate's degree. The majority (74%) had a preschool or regular teaching credential and 35 percent had a Child Development Associate (CDA) credential. Table 5.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 12.7 (Kansas site), 16.8 (Florida site), and 13.7 (New Jersey site). The child-staff ratio was on average 5.5 children to one teacher or program staff person in the Kansas site, 8.4 in the Florida site, and 7.8 in the New Jersey site.

Random Assignment

SFA researchers identified school districts that had SFA and non-SFA elementary schools in the same area. The SFA research team then recruited preschool programs based on whether some of the children from each preschool would transition into both SFA and non-SFA elementary schools. For example, in a district with a

Table 5.2. Primary caregiver demographic characteristics for *Curiosity Corner*

Characteristics	Full sample n = 194	Curriculum comparison	
		Control n = 97	Treatment n = 97
Age at baseline (years), mean	31.1	31.0	31.1
Marital status (%)			
Married	46.4	42.3	50.5
Separated/Divorced	17.0	17.5	16.5
Widowed	‡	‡	‡
Never Married	35.1	39.2	30.9
Race/ethnicity (%)			
White, non-Hispanic	33.7	37.1	30.2
African American, non-Hispanic	47.2	38.1	56.3
Hispanic	14.5	21.6	7.3
Asian or Pacific Islander	2.1	‡	‡
Native American	0.0	0.0	0.0
Multiple/other	2.6	‡	‡
Educational level (%)			
Did not finish high school	17.6	16.5	18.8
High school diploma or GED	32.1	37.1	27.1
Some college	34.2	27.8	40.6
College graduate	16.1	18.6	13.5
Employment (%)			
Full-time	50.5	44.3	56.7
Part-time	13.4	14.4	12.4
Unemployed	35.6	40.2	30.9
Other	‡	‡	0.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

preschool classroom randomly assigned to the treatment condition, it was expected that some of the children from that preschool would attend an SFA elementary school and some of the children would attend a non-SFA elementary school for kindergarten.

Along with the SFA researchers, Mathematic Policy Research, Inc. (MPR) determined the unit of random assignment for each of the three SFA research locations. The MPR research staff randomly assigned schools to curriculum conditions because a school had only one classroom or conditions dictated against varying the curriculum conditions within a school. To increase the precision with which to estimate impacts, MPR staff grouped schools into blocks of two or more and randomly assigned half the preschools in each block to the treatment group and half to the control group. MPR staff formed blocks by matching preschools on easily measured characteristics such as teachers' experience, school location, or score on a state report card system and, in doing so, ensured that those characteristics would be evenly distributed between the overall treatment and control groups. MPR staff used a random number function (RAND function in MS Excel) to generate random numbers. They sorted preschools by block and assigned a random number to each preschool. The preschools were then randomly assigned to treatment and control conditions. The staff assigned the highest-ranking preschool within the block to the treatment condition, the next highest to the control condition, alternating assignment to treatment and control conditions until all preschools were randomly assigned to one of two conditions.

Table 5.3. Preschool teacher characteristics for *Curiosity Corner*

Characteristics	Full sample n = 31	Curriculum comparison	
		Control n = 17	Treatment n = 14
Gender (% female)	97.0	94.0	100.0
Race/ethnicity (%)			
White, non-Hispanic	58.0	82.0	29.0
African American, non-Hispanic	19.0	‡	36.0
Hispanic	13.0	0.0	29.0
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	‡	‡	‡
Educational level (%)			
High school diploma or GED	‡	‡	‡
Associate's degree	16.0	‡	‡
Bachelor's degree	45.0	53.0	36.0
Graduate degree	32.0	29.0	36.0
Current teaching license/certificate (%)	74.0	71.0	79.0
Child Development Associate (CDA) (%)	35.0	24.0	50.0
State-awarded preschool certificate (%)	‡	‡	‡
No credential (%)	‡	‡	‡
Years of teaching experience, overall (mean)	10.3	10.5	10.1
Years of preschool teaching experience (mean)	6.9	7.5	6.2

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Contamination

Because only one classroom from each school or preschool program participated in the evaluation, there was little risk of contamination across the treatment and control conditions.

Control Condition

In the control condition, teachers in the three geographic locations used a variety of curricula. In the Florida site, instruction was primarily based on the *Creative Curriculum* model. In Kansas, teachers used a blended curriculum including *Preschool and Language Stimulation (PALS)* and the *Animated Literacy* (Stone 2002) curriculum models, and teacher-developed curricula. In New Jersey, teachers used a teacher-developed, nonspecific curriculum.

Data Collection

MPR collected the child, parent, teacher, and school data for the SFA sites (New Jersey, Kansas, and Florida) for all three waves of data collection. The fall assessment data collection window for child assessments ranged from October 10, 2003 to November 11, 2003 (New Jersey); September 8, 2003 to November 17, 2003 (Kansas); and October 8, 2003 to November 19, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 35 days in New Jersey, 14 days in Kansas, and 49 days in Florida. The spring pre-kindergarten window was May 10, 2004 to June 19, 2004 (New Jersey); April 5, 2004 to May 17, 2004 (Kansas); and April 5, 2004 to May 7, 2004 (Florida). The kindergarten follow-up window was April 12, 2005 to June 8, 2005 (New Jersey); March 28, 2005 to June 8, 2005 (Kansas); and April 2, 2005 to June 8, 2005 (Florida).

Attrition

Eighteen schools were randomly assigned to treatment and control conditions. All 18 schools and 31 classrooms remained in the study throughout the pre-kindergarten year.

For the child assessment, the baseline (fall 2003) response rate was 98 percent; the spring 2004 response rate was 95 percent; and the kindergarten follow-up response rate was 90 percent.

Implementation

The *Curiosity Corner* curriculum was implemented in 14 treatment classrooms. SFA trainers provided initial training and ongoing support to teachers implementing the curriculum. This support included implementation visits, during which trainers observed teachers' instructional practices and the classroom environment. The trainers provided qualitative feedback during visits that were conducted in the fall, winter, and spring of the preschool year. The SFA curriculum fidelity instrument was used for the implementation visit report. The purpose of the fidelity measure is to determine the current level of implementation of the *Curiosity Corner* curriculum components. A team of five SFA trainers individually visited the *Curiosity Corner* classes at least three times per site beyond the initial training visits and completed implementation visit reports. During these follow-up visits, trainers provided support for teachers' emerging expertise with the program. For example, they identified areas for professional development improvement and addressed teachers' questions and concerns. They also met with administrators to discuss the results of their observations. During these meetings, strengths and areas for improvement were identified. The same team of trainers observed control classes. Using the same implementation visit rating scale used in *Curiosity Corner* classes, they visited each control classroom site at least twice to rate the classes. An initial review of the implementation visit reports and narratives indicated a wide degree of variability in the quality of implementation. Variability was evident among and within sites. The implementation quality appeared to vary by teacher. Some sites had teachers who implemented the curriculum exceptionally well and others very poorly.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from "Not at All" (0) to "High" (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. The *Curiosity Corner* curriculum (2.0) and the control curriculum (1.9) were both rated at the Medium level (2.0) on the global implementation fidelity measure.

Impact Analysis Results

We begin with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) and then present the analyses of the classroom observation data. Our discussion of the results focuses on the combined analyses of the three SFA research sites.

Curiosity Corner—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-5a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-5a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 5.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences between groups on these measures for the fall assessment.

There were no statistically detectable differences between groups on the spring of the pre-kindergarten or kindergarten years on any of these measures.

Based on the analyses of the three mathematics measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences between groups on these measures for the fall assessment.

There was no statistically detectable difference between groups on the TERA for the spring pre-kindergarten assessment. There was, however, a statistically reliable effect for the spring kindergarten assessment ($ES_s = .43, p < .05$).

For the WJ Letter Word Identification test, there was no difference for the spring pre-kindergarten assessment, but there was a statistically reliable difference for the spring kindergarten assessment ($ES_s = .43, p < .05$).

There was no statistically detectable difference for the WJ Spelling test for either the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the three reading measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on pre-reading skills relative to the control condition at the end of pre-kindergarten. However, relative to the control group, results indicate there was a delayed effect of *Curiosity Corner* on reading measures at the end of the kindergarten year.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Curiosity Corner* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences for the fall assessment on either measure.

For the spring of the pre-kindergarten and kindergarten years, there were no statistically detectable differences between groups on either measure.

Based on the analyses of the two language measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There was a statistically significant difference on the SSRS Problem Behaviors scale at the fall assessment ($ES_s = .53, p < .05$); children in the *Curiosity Corner* condition were rated as exhibiting more problem behaviors relative to the control group (follow-up analyses for this finding are included in appendix A). There were no statistically significant differences on the SSRS Social Skills or PLBS for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on the measures of behavioral outcomes.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Curiosity Corner—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-5b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-5b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 5.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically significant difference between groups on the fall observation.

No statistically detectable difference between groups was obtained in spring of the pre-kindergarten year.

Based on the analysis of the ECERS-R, we conclude that *Curiosity Corner* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There was a statistically significant difference at the time of the fall observation on the Arnett Permissiveness scale ($ES_c = -1.46, p < .05$; follow-up analyses for this finding are included in appendix A). Teachers in the *Curiosity Corner* classrooms were rated as being less permissive in their interactions with their students relative to teachers in the control classrooms. (Please see appendix A for additional analyses.) There were no statistically detectable differences on the other scales in fall of the pre-kindergarten year.

For the spring pre-kindergarten assessment, there were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales.

Based on the analyses of the four Arnett scales, we conclude that *Curiosity Corner* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) in the spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences on the Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, or Math Concepts scales. There was a statistically significant difference on the TBRS Book Reading scale ($ES_C = 2.06, p < .001$) indicating that the *Curiosity Corner* teachers provided more book reading activities relative to teachers in the control classrooms.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Curiosity Corner* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Curiosity Corner* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Curiosity Corner* had a positive effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Curiosity Corner* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Curiosity Corner*

The findings for *Curiosity Corner* are summarized in table 5.4.

Table 5.4. Effect sizes for *Curiosity Corner*

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.10	.26	—
CMA-A Mathematics Composite	.01	-.05	—
Shape Composition ¹	.16	.32	—
Reading			
TERA	.10	.43*	—
WJ Letter Word Identification	.09	.43*	—
WJ Spelling	.04	.20	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.18	†	.25
Language			
PPVT	-.01	.14	—
TOLD	-.08	.15	—
Behavior			
SSRS Social Skills	-.06	†	.32
SSRS Problem Behavior ²	.43	†	-.08
PLBS/LBS	-.25	†	.11
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	-.48	—	
Teacher-child interaction			
Arnett Detachment ³	-.41	—	
Arnett Harshness ³	.14	—	
Arnett Permissiveness ³	-.98	—	
Arnett Positive Interactions	.02	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	2.06***	
TBRS Oral Language	†	.37	
TBRS Phonological Awareness	†	.44	
TBRS Print and Letter Knowledge	†	-.99	
TBRS Written Expression	†	-.54	
TBRS Math Concepts	†	-.33	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; *** $p < .001$

¹ Building Blocks, Shape Composition task

² Higher scores on this scale represent more negative child behaviors.

³ Lower scores on this scale represent a more positive classroom environment.

⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Page intentionally left blank.

Chapter 6. *Doors to Discovery* and *Let's Begin with the Letter People*: University of Texas Health Science Center at Houston (Texas site)

Curriculum

The University of Texas Health Science Center at Houston (Texas) researchers implemented the *Doors to Discovery* and *Let's Begin with the Letter People* curricula.

Doors to Discovery

The *Doors to Discovery* curriculum is a pre-kindergarten program that is based on the five areas identified by the International Reading Association and the National Association for the Education of Young Children as the foundation for early literacy success: oral language, phonological awareness, concepts of print, alphabet knowledge and writing, and comprehension.

The program focuses on the use of learning centers and shared literacy activities in the pre-kindergarten classroom. The curriculum is presented in eight thematic units that cover topics such as friendship, communities, nature, society, and health. Classroom practices include teachers' directed activities; large and small group activities; and children's application of skills and independent practice on activities that are tied to the curriculum.

The curriculum components also include family learning activities that are designed to foster partnerships between the school and the family; initial training for teachers and ongoing professional development support; and assessment strategies that are integrated into the curriculum units.

Let's Begin with the Letter People

Like *Doors to Discovery*, *Let's Begin with the Letter People* is a comprehensive pre-kindergarten curriculum that is organized thematically. Literacy learning is integrated across topic areas including science, health and safety, art, mathematics, spatial concepts, and music, as well as development of large and small motor skills. The curriculum focuses on specific literacy and language skills including oral language, phonological and phonemic awareness, and letter knowledge. The curriculum lessons address the development of letter knowledge in multiple contexts (e.g., circle time, small group, large group) and activities (e.g., center activities, story times) that support children's development of language and literacy skills.

The teacher lesson plans incorporate activities from the thematic units that are consistent with the overall *Letter People* curriculum objectives. Classroom practices include teacher directed activities, application of skills, and independent practice on activities that are tied to the curriculum. The physical layout of the *Letter People* classroom includes clearly defined interest centers (e.g., Paint Corner, Block, Drama Center, Mathematics, etc.). The curriculum materials include Letter People (huggables). Each Letter Person represents a letter of the alphabet, and has distinguishing characteristics that is readily associated with the sound represented by that letter.

Sample

The Texas research team recruited Head Start and public pre-kindergarten (Title I and non-Title I) programs for participation in the study. All of the programs were full-day programs. All schools and teachers were recruited before the start of the preschool year. Parental consent was obtained during the first few weeks of the school year. A total of 95 teachers and 625 parents and children were recruited as part of the site-specific study. A subset of 44 teachers/classrooms, and 297 parents and children (101 in *Doors to Discovery* treatment

group, 100 in the *Let's Begin with the Letter People* treatment group, and 96 in the control group) were selected for inclusion in the study sample for the PCER initiative. Data were collected on 293 children and 237 parents at the time of the fall assessment data collection.

In the follow-up year of the evaluation, the sample of schools went from 19 in preschool to 78 schools in kindergarten. The sample of classrooms went from 44 preschool to 149 kindergarten classrooms. The kindergarten sample included 250 children and 264 parents from the original sample of 297 participants. Data were collected on 235 children and 203 parents.

Children and Families

The children were 4.6 years of age at the time of baseline data collection and more than half (55%) were male. The racial/ethnic composition of the sample of children was diverse: 43 percent Hispanic, 30 percent White, and 13 percent African American. Table 6.1 provides additional information on the demographic characteristics of the children in the study sample. There were no statistically detectable differences between the treatment and control groups on these child characteristics.

Table 6.1. Child demographic characteristics for *Doors to Discovery* and *Let's Begin with the Letter People*

Characteristics	Full sample n = 297	Curriculum comparison		
		Control n = 96	Treatment 1 ¹ n = 100	Treatment 2 ² n = 101
Age at baseline (years), mean	4.6	4.7	4.6	4.7
Gender (% male)	54.6	54.3	55.6	54.0
Race/ethnicity (%)				
White, non-Hispanic	30.1	24.1	33.3	32.9
African American, non-Hispanic	13.3	19.3	7.4	12.9
Hispanic	43.0	39.8	49.4	40.0
Asian or Pacific Islander	4.4	7.2	3.7	‡
Native American	0.0	0.0	0.0	0.0
Multiple/other	9.2	9.6	6.2	11.8
Child disability status (parent reported, %)	12.3	16.7	10.1	10.7

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹In Texas, Treatment 1 is *Let's Begin with the Letter People*.

²In Texas, Treatment 2 is *Doors to Discovery*.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 6.2. The average age of the primary caregiver was 34 years. The majority (72%) of the primary caregivers were married. Most reported having had some college (27%) or a bachelor's degree (34%), 17 percent had a high school diploma or GED, and 22 percent had not finished high school. Less than half (40%) of the primary caregivers were employed full-time; 20 percent were employed part-time; and 39 percent were unemployed. There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

Table 6.2. Primary caregiver demographic characteristics for *Doors to Discovery* and *Let's Begin with the Letter People*

Characteristics	Full sample n = 237	Curriculum comparison		
		Control n = 73	Treatment 1 ¹ n = 79	Treatment 2 ² n = 85
Age at baseline (years), mean	34.2	34.6	34.0	34.0
Marital status (%)				
Married	71.7	64.4	75.9	74.1
Separated/Divorced	11.8	13.7	11.4	10.6
Widowed	‡	‡	0.0	‡
Never Married	15.2	20.5	12.7	12.9
Race/ethnicity (%)				
White, non-Hispanic	29.9	21.1	29.1	38.1
African American, non-Hispanic	12.8	18.3	8.9	11.9
Hispanic	43.2	45.1	48.1	36.9
Asian or Pacific Islander	4.7	5.6	5.1	‡
Native American	0.0	0.0	0.0	0.0
Multiple/other	9.4	9.9	8.9	9.5
Educational level (%)				
Did not finish high school	21.8	22.5	19.0	23.8
High school diploma or GED	16.7	15.5	21.5	13.1
Some college	27.4	36.6	25.3	21.4
College graduate	34.2	25.4	34.2	41.7
Employment (%)				
Full-time	39.7	45.2	40.5	34.1
Part-time	19.8	16.4	19.0	23.5
Unemployed	39.2	38.4	36.7	42.4
Other	‡	0.0	‡	0.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹In Texas, Treatment 1 is *Let's Begin with the Letter People*.

²In Texas, Treatment 2 is *Doors to Discovery*.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Teachers

There were 44 teachers who participated in the preschool year intervention study. Most (43 of 44) of the preschool teachers were female, and most were White (55%) or African American (32%). The preschool teachers had on average 14 years of teaching experience, with an average of 8 years of experience teaching preschool. Most of the teachers had a bachelor's (66%) or graduate (14%) degree. Eleven percent of the teachers had an associate's degree, and 9 percent had a high school diploma or GED. The teachers reported having a state-awarded preschool certificate (74%), teaching license or certificate (73%), or a Child Development Associate (CDA) credential (18%). Table 6.3 provides additional information on the characteristics of the preschool sample of teachers. At baseline, the *Doors to Discovery* treatment group teachers had more years of preschool teaching experience relative to the teachers assigned to the *Let's Begin with the Letter People* and control group conditions (10.1 years vs. 8.5 years and 5.8 years, $p < .05$).

Programs/Classrooms

The average preschool class size was 18.6 children. The child-staff ratio was on average 9.1 children to one teacher or program staff person.

Table 6.3. Preschool teacher characteristics for *Doors to Discovery* and *Let's Begin with the Letter People*

Characteristics	Full sample n = 44	Curriculum comparison		
		Control n = 16	Treatment 1 ¹ n = 15	Treatment 2 ² n = 13
Gender (% female)	98.0	100.0	100.0	92.0
Race/ethnicity (%)				
White, non-Hispanic	55.0	50.0	60.0	54.0
African American, non-Hispanic	32.0	38.0	27.0	31.0
Hispanic	‡	‡	‡	‡
Asian or Pacific Islander	‡	‡	0.0	0.0
Native American	0.0	0.0	0.0	0.0
Multiple/other	‡	0.0	‡	‡
Educational level (%)				
High school diploma or GED	9.0	0.0	‡	‡
Associate's degree	11.0	13.0	‡	‡
Bachelor's degree	66.0	69.0	73.0	54.0
Graduate degree	14.0	‡	‡	‡
Current teaching license/certificate (%)	73.0	81.0	67.0	69.0
Child Development Associate (CDA) (%)	18.0	‡	‡	‡
State-awarded preschool certificate (%)	74.0	73.0	73.0	77.0
No credential (%)	‡	‡	0.0	‡
Years of teaching experience, overall (mean)	14.1	11.9	15.2	15.4
Years of preschool teaching experience (mean)	8.0	5.8	8.5	10.1*

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

* $p < .05$

¹ In Texas, Treatment 1 is *Let's Begin with the Letter People*.

² In Texas, Treatment 2 is *Doors to Discovery*.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Random Assignment

Randomization was done during the pilot year of curriculum implementation, and the same assignments were maintained for the second year (2003-04) of implementation as had been used during the pilot study year (2002-03). Most of the teachers were second-year implementers of the *Doors to Discovery* or *Let's Begin with the Letter People* curricula. There were no changes to the group of Head Start teachers and classrooms during the second year of implementation. In the Title I district, one *Let's Begin with the Letter People* teacher was replaced; one *Doors to Discovery* teacher was replaced; and three control group teachers were replaced. In the non-Title I district, one *Let's Begin with the Letter People* teacher was replaced and one control group teacher was replaced.

The Texas research team randomly assigned 76 classrooms using a 3 (Type of Curriculum—*Let's Begin with the Letter People*, *Doors to Discovery*, or Control) x 2 (mentoring versus nonmentoring) design with classrooms from three settings (Head Start, Title I pre-kindergarten, and non-Title I pre-kindergarten). There were 27 control classrooms; 24 *Let's Begin with the Letter People* classrooms; and 25 *Doors to Discovery* classrooms dispersed across the three types of preschool settings. A subset of 45 classrooms was randomly selected for inclusion in the study for the PCER initiative. The 76 preschool teachers were provided with a description of the study and given the option to participate. The names of teachers who consented to participate were included in a hat and 45 classrooms/teachers were randomly selected. One teacher later decided not to participate, and this teacher/classroom was dropped from the sample. The final sample included 44 preschool classrooms. Eight children per classroom were randomly selected for pre- and post-testing from among the larger pool of consented children.

The research team randomly assigned preschools to conditions by first creating a list of all of the preschools and the number of available classrooms in each preschool building. The team then labeled three chips with the names of one of the three conditions (*Doors to Discovery*, *Let's Begin with the Letter People*, and Control) and placed the chips in a box. A chip was randomly pulled from the box and the preschool (and related classrooms) on the list was assigned to that condition. For example, school number three on the list was assigned to the control condition if that chip was pulled from the box when the team got to school number three on their list of schools. This procedure was repeated for different school sites until the target number of classrooms was obtained. All of the preschool classrooms at a preschool site/building were assigned to the same condition. The same procedure that was used to assign schools and classrooms to a condition was used to assign treatment classrooms to the mentoring (mentoring vs. nonmentoring) conditions. The names of teachers in a given treatment condition, within a type of pre-kindergarten setting (e.g., 10 Head Start *Let's Begin with the Letter People* classrooms), were put into a container. Half of the teachers were randomly selected to receive mentoring along with their implementation of the treatment curriculum.

Across the three types of preschool settings, a total of 15 classrooms received training and implemented the *Let's Begin with the Letter People* curriculum and 14 classrooms received training and implemented the *Doors to Discovery* curriculum. Half of the teachers in each treatment curriculum condition were randomly assigned to receive mentoring support on a weekly basis. Twenty-seven classrooms were randomly assigned to a control condition that received no specific curriculum, training, or mentoring. A total of 44 classrooms and 297 children took part in the study.

Contamination

Because all classrooms at a preschool site were assigned to only one of three conditions, there was little risk of contamination across the two treatment and control conditions.

Control Condition

In classrooms in the control condition, teachers used teacher-developed, nonspecific curricula.

Data Collection

RTI International (RTI) collected the child, teacher, and school data for the Texas site for all three waves of data collection. The Texas research team was responsible for conducting the parent interviews in the preschool year. In the kindergarten follow-up year, RTI staff completed the parent interviews. The fall assessment data collection window for child assessments ranged from September 8, 2003 to October 29, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 20 days. The spring pre-kindergarten window was April 15, 2004 to June 11, 2004, and the kindergarten follow-up window was April 8, 2005 to June 29, 2005.

Attrition

Forty-four classrooms were randomly assigned to treatment and control conditions. All 44 classrooms remained in the study throughout the pre-kindergarten year.

For the child assessment, the baseline (fall, 2003) response rate was 99 percent, the spring 2004 response rate was 94 percent, and the kindergarten follow-up response rate was 94 percent.

Implementation

Teachers received curriculum implementation training prior to the start of the 2003-04 school year. The teacher sample included 45 teachers who participated in the pilot year of the study (2002-03), and seven new teachers who started in 2003-04. A total of 44 (37 returning) teachers participated in the study during the second year of implementation. The new teachers received 12 hours and returning teachers received 6 hours of curriculum implementation training.

The research team collected site-specific curriculum fidelity data three times during the preschool year. Control classrooms were not observed using the curriculum-specific fidelity measures. All classrooms were observed using a site-specific measure (the Teacher Behavior Rating Scale) in both treatment and control classrooms in fall and spring of the preschool year.

Implementation Fidelity Ratings

Doors to Discovery

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. The *Doors to Discovery* curriculum was rated Medium (2.13) on the global fidelity measure. The control group classrooms were rated as Low (1.0).

Let's Begin with the Letter People

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. The *Let's Begin with the Letter People* curriculum was rated Medium (1.86) on the global implementation fidelity measure. The control group curriculum was rated Low (1.0) on implementation fidelity level.

Impact Analysis Results

We present analyses for each curriculum separately beginning with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. We first present from the *Doors to Discovery* analyses and then from the *Let's Begin with the Letter People* analyses.

Doors to Discovery—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-6a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-6a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 6.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences between the *Doors to Discovery* group and the control group on any of these measures for the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the three mathematics measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on any of these measures at the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the three reading measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Doors to Discovery* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. At the fall assessment, there was no statistically detectable difference between groups on the PPVT, but there was a statistically reliable difference favoring the *Doors to Discovery* group on the TOLD Grammatical Understanding scale ($ES_s = .38, p < .05$; follow-up analyses for this finding are included in appendix A).

There were no statistically detectable differences on either of these measures for the spring pre-kindergarten or spring kindergarten assessments. Based on the analyses of the two language measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on the three behavioral measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Doors to Discovery—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-6b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-6b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 6.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups on the fall observation.

No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Doors to Discovery* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There was a statistically significant difference at the time of the fall observation on the Arnett Permissiveness scale ($ES_c = 1.06, p < .05$; follow-up analyses for this finding are included in appendix A). Teachers in the *Doors to Discovery* classrooms were rated as being more permissive in their interactions with their students relative to teachers in the control classrooms. There were no statistically detectable differences on the other scales for the fall observation.

There were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interactions scales for the spring pre-kindergarten assessment.

Based on the analyses of the four Arnett scales, we conclude that *Doors to Discovery* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) in spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences on the Written Expression, Phonological Awareness, Oral Language, or Math Concepts scales. There were statistically significant differences on the Book Reading ($ES_c = 1.18, p < .01$) and Print and Letter Knowledge ($ES_c = .90, p < .05$) scales indicating that the *Doors to Discovery* teachers provided more instruction in Book Reading and Print and Letter Knowledge relative to teachers in the control classrooms.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Doors to Discovery* had a positive effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Doors to Discovery* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Doors to Discovery* had a positive effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Doors to Discovery* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Doors to Discovery*

The findings for *Doors to Discovery* are summarized in table 6.4 .

Table 6.4. Effect sizes for *Doors to Discovery*

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.01	-.02	—
CMA-A Mathematics Composite	.13	-.16	—
Shape Composition ¹	-.13	-.12	—
Reading			
TERA	.06	-.05	—
WJ Letter Word Identification	.10	-.09	—
WJ Spelling	.06	-.12	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.18	†	-.09
Language			
PPVT	.15	.18	—
TOLD	.17	.06	—
Behavior			
SSRS Social Skills	-.18	†	-.05
SSRS Problem Behavior ²	-.14	†	.46
PLBS/LBS	-.18	†	-.32
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	.39	—	
Teacher-child interaction			
Arnett Detachment ³	-.07	—	
Arnett Harshness ³	-.38	—	
Arnett Permissiveness ³	.13	—	
Arnett Positive Interactions	.38	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	1.18**	
TBRS Oral Language	†	.59	
TBRS Phonological Awareness	†	.58	
TBRS Print and Letter Knowledge	†	.90*	
TBRS Written Expression	†	.62	
TBRS Math Concepts	†	.37	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$

¹ Building Blocks, Shape Composition task

² Higher scores on this scale represent more negative child behaviors.

³ Lower scores on this scale represent a more positive classroom environment.

⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Let's Begin with the Letter People—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-7a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-7a in appendix D. For all analyses of child-level measures, the following covariates were included (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 6.5.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, CMA-A Composite Score, and Shape Composition). There were no statistically detectable differences on any of these measures at the fall assessment.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on analyses of the three mathematics measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (TERA, WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on any of these measures at the fall assessment.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on analyses of the three reading measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically significant difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically significant differences on these measures at the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on analyses of the two language measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences between groups on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Let's Begin with the Letter People—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-7b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-7b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 6.5.

Overall classroom environment

We conducted a repeated measures analysis on the ECERS-R. There was no statistically detectable difference between groups on the fall observation.

A statistically significant difference between groups was obtained for the spring pre-kindergarten assessment ($ES_c = .82, p < .05$), such that the *Let's Begin with the Letter People* classrooms received higher global classroom quality ratings relative to the control classrooms.

Based on the analysis of the ECERS-R, we conclude that *Let's Begin with the Letter People* had a positive effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There was a statistically significant difference on the Arnett Permissiveness scale on the fall observation ($ES_c = .99, p < .01$; follow-up analyses for this finding are included in appendix A). Teachers in the *Let's Begin with the Letter People* classrooms were rated as being more permissive in their interactions with their students relative to teachers in the control classrooms. There were no statistically detectable differences between groups on the other scales for the fall observation.

No statistically detectable differences were obtained for the spring pre-kindergarten assessment on the Arnett Detachment, Permissiveness, and Positive Interactions scales. There was a statistically significant difference

on the Arnett Harshness scale. *Let's Begin with the Letter People* teachers were rated as being less harsh in their interactions with their students relative to teachers in control classrooms ($ES_C = -.95, p < .05$).

Based on the analyses of the four Arnett scales, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (TBRS Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) in spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences on the Book Reading, Written Expression, Phonological Awareness, Oral Language, or Math Concepts scales. There was a statistically significant difference on the Print and Letter Knowledge scale ($ES_C = .99, p < .05$) indicating that the *Let's Begin with the Letter People* teachers provided more instruction on print and letter knowledge relative to teachers in the control classrooms.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Let's Begin with the Letter People* had a positive effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analyses of the TBRS Book Reading and Oral Language scales, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Let's Begin with the Letter People*

The findings for *Let's Begin with the Letter People* are summarized in table 6.5.

Table 6.5. Effect sizes for *Let's Begin with the Letter People*

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	-.10	-.13	—
CMA-A Mathematics Composite	.15	-.07	—
Shape Composition ¹	.21	-.06	—
Reading			
TERA	.02	-.13	—
WJ Letter Word Identification	.10	-.18	—
WJ Spelling	.17	-.06	—
Phonological awareness			
Pre-CTOPPP/CTOPP	-.13	†	-.13
Language			
PPVT	-.03	.00	—
TOLD	.08	-.12	—
Behavior			
SSRS Social Skills	-.27	†	.24
SSRS Problem Behavior ²	-.06	†	.06
PLBS/LBS	-.44	†	-.10
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	.82*	—	
Teacher-child interaction			
Arnett Detachment ³	-.07	—	
Arnett Harshness ³	-.95*	—	
Arnett Permissiveness ³	-.05	—	
Arnett Positive Interactions	.48	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	.63	
TBRS Oral Language	†	.44	
TBRS Phonological Awareness	†	.66	
TBRS Print and Letter Knowledge	†	.99*	
TBRS Written Expression	†	.60	
TBRS Math Concepts	†	.24	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹ Building Blocks, Shape Composition task

² Higher scores on this scale represent more negative child behaviors.

³ Lower scores on this scale represent a more positive classroom environment.

⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Chapter 7. *Early Literacy and Learning Model (ELLM)*: University of North Florida (Florida-UNF site)

Curriculum

The University of North Florida (Florida-UNF) team implemented *Early Literacy and Learning Model (ELLM)*, a literacy-focused curriculum and support system designed for young children from low-income families. The *ELLM* program components include the following:

- curriculum and literacy building blocks;
- assessment for instructional improvement;
- professional development for literacy coaches and teachers;
- family involvement; and
- collaborative partnerships.

The *ELLM* curriculum and support system is designed to enhance existing classroom curricula by specifically focusing on children's early literacy skills and knowledge.

The *ELLM* curriculum materials include a set of literacy performance standards; monthly literacy packets; targeted instructional strategies; resource guides for teachers; a book lending library; family and teacher tip sheets; and literacy calendars. One hour of daily literacy instruction is required to implement the *ELLM* literacy building blocks. Trained literacy coaches provide instructional support to preschool teachers who use the curriculum.

The *ELLM* program contains a family involvement action plan. Families have access to many resources, including a classroom book-lending library that enables children to take books home daily to share with their parents. Parents receive monthly family tip sheets and calendars with suggestions for literacy activities they can engage in with their children. Parents also have the opportunity to engage in preschool site-based family activities during the school year.

As part of the Florida-UNF complementary study, the *ELLM* program included two evaluation instruments: the *Test of Early Reading Ability-Third Edition, Form A (TERA-3)*³, and the *Alphabet Letter Recognition Inventory (ALRI)* were used as assessment tools. *ELLM* teachers used results from these assessments to identify children's literacy needs and inform classroom literacy instruction. For example, children's fall scores were used to help teachers focus instruction and identify children for targeted instruction in phonological awareness and letter recognition.

Sample

During the 2003-04 academic year, the Florida-UNF research team recruited 28 preschool classrooms from three geographic locations in Florida. The sampled classrooms included Head Start, subsidized, faith-based, and early intervention pre-kindergarten classrooms. All of the classrooms were full-day programs. Twenty-eight classrooms and teachers were recruited to participate in the study. The Florida-UNF research team attended site orientation and/or parent meetings to recruit participants. The teachers and program administrators assisted with the recruitment efforts. Consent forms were sent home to parents and teachers

³ The TERA used in the *ELLM* classrooms was a different version (Form A) than that used for the PCER evaluation study assessment (Form B).

collected signed consent forms from parents. A total of 297 children and 294 parents were recruited for participation in the study. The final sample included 244 children (137 treatment, 107 control) and 243 parents. Data were collected on 243 children and 204 parents at the time of the fall assessment data collection. During the study year, two sites (one control, one intervention) withdrew from the study.

In the follow-up year of the evaluation (the 2004-05 academic year), the sample of schools went from 28 preschools to 119 schools with kindergarten classrooms. The sample of classrooms went from 28 preschool classrooms to 175 kindergarten classrooms. The kindergarten sample included 237 children and 236 parents from original sample of 248 participants. Data were collected on 218 children and 177 parents.

Children and Families

The average age of children was 4.6 years at the time of fall assessment data collection and half (50%) was male. The overall sample was primarily African American (71%) with smaller percentages of White (14%) and Hispanic (8%) children. Table 7.1 provides additional information on the demographic characteristics of the children in the study sample. There were no statistically detectable differences between the treatment and control groups on these child characteristics.

Table 7.1. Child demographic characteristics for *Early Literacy and Learning Model*

Characteristics	Full sample n = 244	Curriculum comparison	
		Control n = 107	Treatment n = 137
Age at baseline (years), mean	4.6	4.6	4.6
Gender (% male)	50.0	48.6	51.1
Race/ethnicity (%)			
White, non-Hispanic	13.8	17.0	11.3
African American, non-Hispanic	71.1	69.1	72.6
Hispanic	7.8	4.3	10.5
Asian or Pacific Islander	‡	‡	‡
Native American	0.0	0.0	0.0
Multiple/other	6.0	7.4	4.8
Child disability status (parent reported, %)	12.7	8.9	15.8

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 7.2. The average age of the primary caregiver was 31 years. Almost half (40%) of the primary caregivers were never married; 37 percent were married at the time of the fall assessment data collection. More than one-third of the primary caregivers reported having had some college (36%) or had graduated from college (6%); 37 percent had a high school diploma or GED; and 22 percent had not finished high school. More than half (54%) of the primary caregivers were employed full-time; 11 percent were employed part-time; and 33 percent were unemployed. At baseline, a higher percentage of parents in the treatment group had completed some post-high school education relative to those assigned to the control group (41% vs. 29%, $p < .01$).

Teachers

There were 28 teachers who participated in the preschool year intervention study. All were female. The majority identified themselves as African American (64%) or White (21%). The preschool teachers had on

Table 7.2. Primary caregiver demographic characteristics for *Early Literacy and Learning Model*

Characteristics	Full sample n = 204	Curriculum comparison	
		Control n = 90	Treatment n = 114
Age at baseline (years), mean	30.9	31.0	30.8
Marital status (%)			
Married	36.8	41.1	33.3
Separated/Divorced	20.6	18.9	21.9
Widowed	2.5	‡	‡
Never Married	40.2	36.7	43.0
Race/ethnicity (%)			
White, non-Hispanic	13.7	16.7	11.4
African American, non-Hispanic	74.0	72.2	75.4
Hispanic	6.4	‡	9.6
Asian or Pacific Islander	2.0	‡	‡
Native American	0.0	0.0	0.0
Multiple/other	3.9	6.7	‡
Educational level (%)			
Did not finish high school	21.6	32.2	13.2
High school diploma or GED	36.8	28.9	43.0
Some college	35.8	28.9	41.2**
College graduate	5.9	10.0	‡
Employment (%)			
Full-time	54.4	45.6	61.4
Part-time	10.8	12.2	9.6
Unemployed	33.3	38.9	28.9
Other	‡	‡	0.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

** $p < .01$

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

average 11 years of teaching experience, with an average of seven years teaching preschool. Fifty percent of the teachers had a high school diploma or GED and 21 percent had a bachelor's degree. Many of the teachers reported having a state-awarded preschool certificate (52%); a teaching license or certificate (46%); or a Child Development Associate (CDA) credential (46%). Eighteen percent reported having no teacher certification credentials. Table 7.3 provides additional information on the characteristics of the preschool sample of teachers. At baseline, teachers in the treatment group had more years of experience teaching in a preschool setting relative to those assigned to the control group (9 years vs. 4 years, $p < .01$).

Programs/Classrooms

The average preschool class size was 15.5 children. The child-staff ratio was an average of 9.6 children to one teacher or program staff person.

Random Assignment

Randomization was done during the pilot-year study (2002-03). The original random assignment procedure and changes that were made during the evaluation study year are summarized here. A total of 30 classrooms and teachers were included in the pilot-year study sample. Preschool classrooms were randomly assigned to

Table 7.3. Preschool teacher characteristics for Early Literacy and Learning Model

Characteristics	Full sample n = 28	Curriculum comparison	
		Control n = 14	Treatment n = 14
Gender (% female)	100.0	100.0	100.0
Race/ethnicity (%)			
White, non-Hispanic	21.0	29.0	‡
African American, non-Hispanic	64.0	50.0	79.0
Hispanic	‡	‡	‡
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	‡	‡	0.0
Educational level (%)			
High school diploma or GED	50.0	50.0	50.0
Associate's degree	‡	‡	‡
Bachelor's degree	21.0	29.0	‡
Graduate degree	‡	‡	‡
Current teaching license/certificate (%)	46.0	43.0	50.0
Child Development Associate (CDA) (%)	46.0	43.0	50.0
State-awarded preschool certificate (%)	52.0	62.0	43.0
No credential (%)	18.0	‡	‡
Years of teaching experience, overall (mean)	10.7	9.1	12.3
Years of preschool teaching experience (mean)	6.7	4.0	9.4**

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

** $p < .01$

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

treatment or control conditions. The Florida-UNF researchers recruited preschool programs from three distinct geographic locations within the state. The research team first identified elementary school neighborhoods in each geographic location (Counties A, B, and C) with low-performing schools. Using the Florida Department of Education's school grading report card system,⁴ the research team identified grade D and F elementary schools in each of the three counties. It was expected that children from the preschool programs in these low-performing elementary school neighborhoods would transition into these grade D and F elementary schools during the kindergarten year of the study. Preschool programs within the low-performing elementary school neighborhoods were randomly selected for inclusion in the sampling pool of preschool programs.

The sampled preschool classrooms included Head Start, subsidized, faith-based, and early intervention pre-kindergarten programs. Thirty preschool classrooms (10 in County A, 10 in County B, and 10 in County C) were randomly assigned to the treatment or control condition. Only one preschool classroom per preschool was randomly assigned to the treatment or control condition.

⁴ All schools in Florida receive a grade based on the following: (1) percentage of students meeting high standards of the Florida Comprehensive Assessment Test (FCAT)-achievement scores of Level 3 or above; (2) percentage of students making learning gains; and (3) adequate progress of the lowest 25 percent of the students in the school. Each school receives a certain number of points for each of three categories. The points are summed to create a total score. The total score is converted into a letter: grade A (410 points or more), grade B (380 to 409 points), grade C (320 to 379), grade D (280 to 319), and grade F (less than 250). For a grade of A, 95 percent were tested and at least 50 percent of the lowest readers must have made gains in the current school year. For a grade of B or C at least 50 percent of the lowest readers must have made gains in one or two consecutive years. Information reported here is based on the school grade categories in use during the 2003-04 school year. Source: <http://schoolgrades.fldoc.org>.

All of the preschool classrooms in a given elementary school neighborhood were randomly assigned to only one of two conditions (*ELLM* or control). In County A, the research team used a random number software program to randomly assign preschools to conditions. In Counties B and C, all preschools were identified, preschool names were written on strips of paper, and placed in a hat. Preschool classrooms were randomly assigned, one at a time, first to treatment and then to control, until all of the preschools were assigned to one of two conditions.

During the evaluation study year, 14 of the 15 pilot-year treatment classrooms remained in the study. To replace a preschool program that withdrew from the study during the pilot year, a classroom was selected from the site-specific classrooms randomly assigned to implement *ELLM* during the pilot year. Fifteen new control classroom teachers were recruited in the second year of the study to replace those from the pilot year who then received *ELLM* training during the second year as part of their agreement to participate in the pilot study. Preschool programs located within the original elementary school neighborhoods were identified and new control classroom sites were randomly selected from a pool of preschool classrooms in each elementary school neighborhood. Fifteen new control group teachers participated in the study during the intervention year (2003-04). The final evaluation study sample included a total of 28 classrooms (28 of the 30 classrooms remained in the study for the duration of the pre-kindergarten school year) and 299 children.

Contamination

Because all preschool classrooms were assigned to only one of two conditions, there was little risk of contamination across the treatment and control conditions.

Control Condition

A number of curricula were represented in the control classrooms including *Creative Curriculum* (Dodge, Colker, and Heroam 2002), *Beyond Centers and Circletime* (Phelps 2002), *High Reach Learning Pre-K* (High Reach Learning 1997a and 1997b), and *High/Scope* (Hohmann and Weikart 2002).

Data Collection

RTI International (RTI) collected the child, teacher, and school data for the three Florida-UNF sites (Counties A, B, and C) for all three waves of data collection. The Florida-UNF research team was responsible for conducting the parent interviews in the preschool year. In the kindergarten follow-up year, RTI staff completed the parent interviews. The fall assessment data collection window for child assessments ranged from September 8, 2003 to October 30, 2003 (County B); September 4, 2003 to October 22, 2003 (County A); and September 15, 2003 to December 4, 2003 (County C). The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 27 days in County B, 28 days in County A, and 21 days in County C. The spring pre-kindergarten window was April 2, 2004 to May 7, 2004 (County B); April 13, 2004 to May 6, 2004 (County A); and May 3, 2004 to June 30, 2004 (County C). The kindergarten follow-up window was April 5, 2005 to June 27, 2005 (County B); April 4, 2005 to June 22, 2005 (County A); and April 8, 2005 to June 15, 2005 (County C).

Attrition

Thirty classrooms were randomly assigned to treatment and control conditions. The final sample included 28 classrooms that remained in the study throughout the pre-kindergarten year.

For the child assessment, the fall assessment response rate was 98 percent, the spring 2004 response rate was 92 percent, and the kindergarten follow-up response rate was 92 percent.

Implementation

Eleven of the 14 teachers in the *ELLM* condition were in their second year of implementation of the curriculum at the time of the evaluation.

The *ELLM* literacy curriculum was implemented in combination with the existing comprehensive curricula. Three *ELLM* literacy coaches were trained during a 5-day training session in August 2003. A 2-day follow-up training institute was held in October 2003. Ongoing training of coaches included weekly local seminars at each site location, monthly regional seminars, and monthly regional collaboration team meetings. Teacher training included a 2-day summer training session; weekly classroom visits by *ELLM* literacy coaches; monthly site-specific literacy team meetings; and quarterly teacher get-togethers. Teacher training focused on the *ELLM* curriculum, *ELLM* learning materials, and strategies to help children acquire important emergent literacy skills.

The *ELLM* literacy coaches made weekly literacy visits (1 hour) to intervention classrooms. *ELLM* literacy coaches hosted monthly literacy team meetings at each site location. At the monthly meetings, the literacy coaches distributed monthly materials and resources; demonstrated the use of monthly literacy packets and children's books; shared instructional ideas, and highlighted targeted activities. The teachers also gave the coaches feedback on the effectiveness of their classroom visits and how to better meet the needs of individual teachers.

The Florida-UNF research team collected videotaped data to measure the fidelity of *ELLM* curriculum implementation. Trained videographers videotaped teachers twice (fall 2003 and spring 2004) during the school year. The videotapes were segmented and coded to analyze fidelity of implementation, and were coded to capture the presence or absence of the critical *ELLM* elements in the intervention classrooms. The possible scores on the *ELLM* fidelity-of-use instrument ranged from 0 to 147. A high level of *ELLM* curriculum implementation is defined as 80 percent (118) of possible points on the fidelity-of-use instrument. This level of implementation is aligned with the competent level on the *ELLM* teacher implementation measure. A low level of *ELLM* implementation is reflected by 60 percent (0-88) of possible points on the fidelity-of-use instrument.

Site-Specific Fidelity Ratings

On the site-specific fidelity measure across both assessment times, the intervention classrooms were rated at a Low or Medium level of implementation. No intervention teacher was rated as a high implementer. With one exception, the control classrooms were rated at a Low level of implementation. One control classroom received a Medium level of implementation rating during the spring 2004 observation.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from "Not at All" (0) to "High" (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. The *ELLM* curriculum was rated at the Medium implementation fidelity level (2.5). The research team did not provide the RTI evaluation staff with a global fidelity rating (using the four-point scale) for the control group classrooms at their research site.

Impact Analysis Results

We begin with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) and then present the analyses of the classroom observation data. Our discussion of the results focuses on the combined analysis of the three sites.

Early Literacy and Learning Model—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-8a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-8a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ES_s) are presented in table 7.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences at the fall assessment.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten or kindergarten assessments.

Based on the analyses for the three mathematics measures, we conclude that *ELLM* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten or kindergarten assessments.

Based on the analyses for the three reading measures, we conclude that *ELLM* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically significant difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *ELLM* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on either measure for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically significant mean differences between groups on either measure. However, for the spring kindergarten assessment, there were statistically reliable differences on the PPVT ($ES_s = .34, p < .05$) and the TOLD Grammatical Understanding ($ES_s = .44, p < .05$).

Based on the analyses of the two language measures, we conclude that *ELLM* did not have a statistically detectable effect on language development relative to the control condition in the pre-kindergarten year. However, results indicate there was a delayed effect of *ELLM* on language development relative to the control condition at the end of the kindergarten year.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on the behavioral measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *ELLM* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Early Literacy and Learning Model—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-8b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-8b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 7.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups for the fall observation. No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *ELLM* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures in fall of the pre-kindergarten year.

There were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales for the spring pre-kindergarten observation.

Based on the analyses of the four Arnett scales, we conclude that *ELLM* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) in spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically reliable differences between groups on any of the TBRS scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *ELLM* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *ELLM* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *ELLM* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *ELLM* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Early Literacy and Learning Model*

The findings for *ELLM* are summarized in table 7.4.

Table 7.4. Effect sizes for Early Literacy and Learning Model

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.10	.26	—
CMA-A Mathematics Composite	.01	-.05	—
Shape Composition ¹	-.14	.03	—
Reading			
TERA	.15	.30	—
WJ Letter Word Identification	-.05	.00	—
WJ Spelling	.11	.04	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.18	†	.08
Language			
PPVT	.17	.34*	—
TOLD	.15	.44**	—
Behavior			
SSRS Social Skills	-.06	†	.27
SSRS Problem Behavior ²	-.24	†	.23
PLBS/LBS	.14	†	.04
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	-.48	—	
Teacher-child interaction			
Arnett Detachment ³	-.41	—	
Arnett Harshness ³	-.40	—	
Arnett Permissiveness ³	-.24	—	
Arnett Positive Interactions	.29	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	.32	
TBRS Oral Language	†	.14	
TBRS Phonological Awareness	†	.53	
TBRS Print and Letter Knowledge	†	.41	
TBRS Written Expression	†	-.22	
TBRS Math Concepts	†	-.92	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$ ¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Chapter 8. *Language-Focused Curriculum*: University of Virginia (Virginia site)

Curriculum

The University of Virginia (Virginia) research team evaluated the *Language-Focused Curriculum (LFC)*. The *LFC* was developed through a 1985 Model Demonstration Project funded by the U.S. Department of Education to the University of Kansas to design a Language Acquisition Preschool. The *LFC* was designed for use with 3- to 5-year-old children with language limitations, including children with language impairment; children from disadvantaged backgrounds; and English-language learners.

The curriculum components include the following:

- thematic organization of content by day, week, and month;
- use of daily dramatic play to teach and use new linguistic concepts;
- use of both teacher-led and child-led activities to organize daily experiences;
- explicit attention to oral language goals across the day; and
- teacher use of the eight key “language stimulation techniques” when interacting with children in the classroom.

The *LFC* emphasizes the daily inclusion of high-quality teacher-child conversations within teacher-led and child-led interactions.

Sample

The Virginia research team recruited 14 teachers and preschool classrooms to participate in the study. A combination of Head Start and public pre-kindergarten classrooms was recruited. All of the programs were full-day programs. Teachers received incentives for participating in the study. Teachers and school administrators assisted with the recruitment of parents and children. The parent and child recruitment process occurred during the first few weeks of the school year. An incentive (storybooks) was offered to children as part of the parental consenting process. A total sample of 205 children and parents were recruited for the study. The average parental consent rate was 94 percent (95% for the treatment group, 93% for the control group). The final sample included 195 children (97 treatment, 98 control) and parents. Data were collected on 182 children and 181 parents at the time of the fall baseline data collection.

In the follow-up year of the evaluation, the sample of schools went from five in pre-kindergarten to 21 schools in kindergarten. The sample of classrooms went from 14 preschool to 54 kindergarten classrooms. The kindergarten sample included 189 of the original sample of 195 children. Data were collected on 189 children and 174 parents.

Children and Families

The children were 4.6 years of age at the time of baseline data collection and slightly more than half (53%) were male. The majority of the sample of preschoolers were White (71%) or African American (21%). Table 8.1 provides additional information on the demographic characteristics of the children in the Virginia study sample. There were no statistically detectable differences between the treatment and control groups on these child characteristics.

Table 8.1. Child demographic characteristics for *Language-Focused Curriculum*

Characteristics	Full sample n = 195	Curriculum comparison	
		Control n = 98	Treatment n = 97
Age at baseline (years), mean	4.6	4.6	4.6
Gender (% male)	52.7	52.7	52.8
Race/ethnicity (%)			
White, non-Hispanic	70.8	67.4	74.4
African American, non-Hispanic	20.8	25.0	16.3
Hispanic	4.5	5.4	‡
Asian or Pacific Islander	0.0	0.0	0.0
Native American	‡	0.0	‡
Multiple/other	2.8	‡	‡
Child disability status (parent reported, %)	17.7	16.1	19.3

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 8.2. The average age of the primary caregiver was 30 years. More than half (54%) of the primary caregivers were married, and 22 percent were never married. Almost half (45%) reported having a high school diploma or GED; 20 percent had not finished high school; 29 percent had some college education; and 7 percent had a BA. Less than half (46%) of the primary caregivers were employed full-time, 39 percent were unemployed, and 14 percent were employed part-time. There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

Teachers

There were 14 teachers who participated in the preschool-year intervention study. All of the preschool teachers were female, and all were White. On average, the preschool teachers had 11 years of teaching experience, with an average of 8 years of experience teaching preschool. The majority of teachers had a bachelor's (71%) degree. The majority of teachers reported having a state teacher certificate (71%). Table 8.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 13 children. The child-staff ratio was on average 6.3 children to one teacher or program staff person.

Random Assignment

The research team identified and recruited a convenience sample of preschools from two counties in Virginia (one rural county and one suburban county). Along with the Virginia researchers, Mathematica Policy Research, Inc. (MPR) determined the unit of random assignment for this research site. The MPR research staff randomly assigned individual classrooms to conditions after it was determined that the experimental curriculum could be introduced in one classroom without affecting neighboring classrooms in the same school and, second, that preschool staff were willing to use different curricula in the same setting. Individual classrooms within schools were randomly assigned to treatment and control conditions. To increase the

Table 8.2. Primary caregiver demographic characteristics for *Language-Focused Curriculum*

Characteristics	Full sample n = 179	Curriculum comparison	
		Control n = 93	Treatment n = 86
Age at baseline (years), mean	29.8	30.5	29.2
Marital status (%)			
Married	54.2	50.5	58.1
Separated/Divorced	22.3	22.6	22.1
Widowed	‡	‡	0.0
Never Married	22.3	24.7	19.8
Race/ethnicity (%)			
White, non-Hispanic	74.9	76.3	73.3
African American, non-Hispanic	18.4	20.4	16.3
Hispanic	5.0	‡	7.0
Asian or Pacific Islander	0.0	0.0	0.0
Native American	‡	0.0	‡
Multiple/other	‡	0.0	‡
Educational level (%)			
Did not finish high school	19.6	22.6	16.3
High school diploma or GED	44.7	40.9	48.8
Some college	29.1	31.2	26.7
College graduate	6.7	5.4	8.1
Employment (%)			
Full-time	46.4	47.3	45.3
Part-time	14.0	11.8	16.3
Unemployed	39.1	39.8	38.4
Other	‡	‡	0.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

precision with which to estimate impacts, MPR grouped classrooms into blocks of two and randomly assigned half the classrooms in each block to the treatment group and half to the control group. The MPR research staff formed blocks by matching classrooms on easily measured characteristics such as teachers' experience, school location, or score on a state report card system and, in doing so, increased the probability that those characteristics would be evenly distributed between the overall treatment and control groups. MPR staff used a random number function (RAND function in MS Excel) to generate random numbers. They sorted the classrooms by block and assigned a random number to each classroom. The classrooms were then randomly assigned to treatment and control conditions. The staff assigned the highest-ranking classroom within the block to the treatment condition, the next highest to the control condition, alternating assignment to treatment and control conditions until all classrooms were randomly assigned to one of two conditions. A total of 14 classrooms (7 treatment and 7 control) were randomly assigned to conditions.

Table 8.3. Preschool teacher characteristics for *Language-Focused Curriculum*

Characteristics	Full sample n = 14	Curriculum comparison	
		Control n = 7	Treatment n = 7
Gender (% female)	100.0	100.0	100.0
Race/ethnicity (%)			
White, non-Hispanic	100.0	100.0	100.0
African American, non-Hispanic	0.0	0.0	0.0
Hispanic	0.0	0.0	0.0
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	0.0	0.0	0.0
Educational level (%)			
High school diploma or GED	‡	‡	‡
Associate's degree	‡	‡	0.0
Bachelor's degree	71.0	57.0	86.0
Graduate degree	‡	‡	0.0
Current teaching license/certificate (%)	71.0	71.0	71.0
Child Development Associate (CDA) (%)	‡	‡	‡
State-awarded preschool certificate (%)	29.0	‡	‡
No credential (%)	‡	‡	‡
Years of teaching experience, overall (mean)	11.4	11.4	11.3
Years of preschool teaching experience (mean)	8.0	7.4	8.6

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Contamination

In each of the five participating schools, there were both treatment and control classrooms. To reduce the possibility of contamination across conditions, the researchers monitored the classrooms to ensure that treatment group teachers were not sharing materials and instructional practices with the control group teachers.

Control Condition

In the control condition, the teachers reported using *High/Scope* curriculum materials, but the extent of *High/Scope* curriculum implementation in the control classrooms was not formally assessed.

Data Collection

MPR collected the child, parent, teacher, and school data for the Virginia site for all three waves of data collection. The fall assessment data collection window for child assessments ranged from September 29, 2003 to November 11, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 28 days. The spring pre-kindergarten window was April 1, 2004 to June 18, 2004, and the kindergarten follow-up window was March 29, 2005 to June 8, 2005.

Attrition

Fourteen classrooms were randomly assigned to treatment or control condition. All 14 classrooms remained in the study from the beginning of the pre-kindergarten year through the spring of the pre-kindergarten year.

For the child assessment, the baseline (fall 2003) response rate was 93 percent, the spring 2004 response rate was 96 percent, and the kindergarten follow-up response rate was 97 percent.

Implementation

Seven classrooms were assigned to implement *LFC* and seven classrooms maintained the prevailing curriculum (*High/Scope*). Five of the seven teachers and seven teaching assistants completed a 3-day training workshop on *LFC* implementation in August 2003. The workshop content included background information on language development. A one-on-one make-up training session was provided to the remaining two teachers who could not attend the initial workshop. Additional follow-up training sessions were held in November 2003, and January/February 2004. In November 2003 treatment group teachers attended an informal on-site 2-hour workshop to discuss teachers' concerns with the implementation of the *LFC*, review the feedback from the first round of classroom observations, and review language stimulation techniques and appropriate use. All of the teachers attended a 3-hour workshop in January/February 2004. The workshop topic was *Being a Conversational Partner*, which focused on language stimulation in the *LFC*, with periodic follow-up training sessions for further discussion and description of implementation activities. All teachers maintained professional development logs throughout the school year to evaluate the extent of professional development experienced by treatment and control group teachers. Site-specific curriculum fidelity observations were conducted in treatment and control classrooms in the fall and spring of the preschool year.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. Both the *LFC* and the control group curriculum were rated at the Medium (2.0) level on the global implementation fidelity measure.

Impact Analysis Results

We begin with the analyses of the child outcomes (i.e., mathematics, reading, phonological awareness, and language assessments) and then present the analyses of the classroom observation data.

Language-Focused Curriculum—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-9a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-9a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 8.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences between groups on the spring pre-kindergarten or kindergarten assessments.

Based on the analyses for the three mathematics measures, we conclude that the *LFC* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three reading measures, we conclude that the *LFC* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference between groups on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that the *LFC* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten and spring kindergarten assessments.

Based on the analyses for the two language measures, we conclude that the *LFC* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the

pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that the *LFC* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Language-Focused Curriculum—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-9b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-9b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

The results from the analysis of the overall classroom environment Early Childhood Environment Rating Scale-Revised (ECERS-R) and teacher-child relationships (Arnett measure) for the Virginia site are not included in this report because of data integrity concerns. During the baseline data collection, one observer completed the observational ratings in 8 of the 12 classrooms at this research site. It was later determined that the ECERS-R and Arnett ratings from these eight classrooms were inflated. Due to concerns regarding the integrity of the data from these eight classrooms, the decision was made to exclude the classroom quality and teacher-child relationships data for this site from the report.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) in spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted; the covariates were (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_C) are presented in table 8.4.

There were no statistically detectable differences between groups on the TBRS scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that the *LFC* did not have statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that the *LFC* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that the *LFC* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that the *LFC* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for Language-Focused Curriculum

The findings for *LFC* are summarized in table 8.4.

Table 8.4. Effect sizes for Language-Focused Curriculum

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.20	.11	—
CMA-A Mathematics Composite	.08	.00	—
Shape Composition ¹	.08	.06	—
Reading			
TERA	.16	.05	—
WJ Letter Word Identification	.11	.02	—
WJ Spelling	.25	.11	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.20	†	.03
Language			
PPVT	.02	-.09	—
TOLD	.01	-.07	—
Behavior			
SSRS Social Skills	-.42	†	-.07
SSRS Problem Behavior ²	.37	†	-.05
PLBS/LBS	-.27	†	.10
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	—	—	
Teacher-child interaction			
Arnett Detachment ³	—	—	
Arnett Harshness ³	—	—	
Arnett Permissiveness ³	—	—	
Arnett Positive Interactions	—	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	-.79	
TBRS Oral Language	†	.87	
TBRS Phonological Awareness	†	.92	
TBRS Print and Letter Knowledge	†	.33	
TBRS Written Expression	†	.99	
TBRS Math Concepts	†	.20	

— Not available. Data were collected but not reported.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

¹ Building Blocks, Shape Composition task

² Higher scores on this scale represent more negative child behaviors.

³ Lower scores on this scale represent a more positive classroom environment.

⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Chapter 9. Literacy Express and DLM Early Childhood Express supplemented with Open Court Reading Pre-K: Florida State University (Florida-FSU site)

Curriculum

The Florida State University (Florida-FSU) research team chose to evaluate two curricula: *Literacy Express* and *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*.

Literacy Express

The Florida-FSU research team implemented the *Literacy Express* curriculum. *Literacy Express* is a preschool literacy-focused curriculum that is designed to promote children's emergent literacy skills. The curriculum is structured around thematic units. The units, and the games and activities within each unit, are sequenced in order of complexity. Each thematic unit of the curriculum includes selected children's books that address theme-relevant vocabulary for small- and large-group reading activities. In addition, each thematic unit includes small-group activities that provide children with the opportunity to attend to and practice the skills needed to develop oral language, phonological sensitivity, and print awareness, and to receive individual feedback needed to master each developmental level. Small-group activities are conducted 3-4 times a week. The curriculum provides guidance to teachers on grouping children who are progressing at similar rates. The large-group and extension activities provide opportunities for children to use new skills in novel and varied contexts.

DLM Early Childhood Express supplemented with Open Court Reading Pre-K

The Florida-FSU research team implemented the Open Court literacy-focused curriculum in conjunction with *DLM Early Childhood Express* comprehensive curriculum. The *Open Court Reading Pre-K* curriculum is a literacy-focused curriculum. The curriculum content is presented in eight thematic units that address children's identity, families, friends, social interactions, transportation, the physical senses, nature, and transitions. Phonological, phonemic, and print-awareness activities are incorporated into each lesson. Comprehension activities are also included in each lesson to help promote children's understanding of literature. Each day, teachers read literature selections that focus on the topic that is in a thematic unit. The curriculum includes a home component to encourage home/school connections by providing parents with suggestions for activities that they can engage in at home with their children.

The *DLM Early Childhood Express* Program is a comprehensive curriculum. The *DLM Early Childhood Express* curriculum is designed to promote children's social, emotional, intellectual, aesthetic, and physical development through the use of hands-on learning experiences. The curriculum has 36 weekly themes that address the following content areas: literacy, mathematics, science, social studies, fine arts, health/safety, personal/social development, physical movement, and technology. Each thematic unit includes more than 200 age-appropriate, hands-on learning activities that are designed to promote children's social, emotional, intellectual, aesthetic, and physical development.

By integrating the research-based instruction from *Open Court Reading Pre-K* with the comprehensive instructional framework of *DLM Early Childhood Express*, children received instruction that is intended to provide them with a strong foundation in oral language and print awareness as well as research-based instruction in phonics and early decoding and comprehension skills.

Sample

The description of the recruitment process applies to both curricula. The research team recruited public pre-kindergarten programs for participation in the study. Principals from elementary schools were provided information regarding the proposed project and invited to participate. Two teachers from each of the 16 participating schools were recruited to participate in the study. All of the programs were full-day programs. No incentives were offered to teachers. The final study sample included 30 teachers and classrooms across three conditions (9 control, 10 *Literacy Express*, and 11 *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*). Teachers assisted with the recruitment of parents and children to participate in the study. The parental consent process began at the beginning of the school year and continued into the first few weeks of school. The average parental consent rate was 94 percent (95% for the treatment group; 93 percent for the control group). A total of 297 children (99 in the *Literacy Express* treatment group; 101 in the *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K* treatment group; and 97 in the control group) and parents were recruited. Data were collected on a total of 282 children and 270 parents at the time of the fall baseline data collection.

In the follow-up year of the evaluation, the sample of schools went from 17 in pre-kindergarten to 46 schools in kindergarten. The sample of classrooms went from 30 preschool to 145 kindergarten classrooms. Data were collected on 237 children and 223 parents from the original sample.

Children and Families

The children were 4.6 years of age at the time of baseline data collection and slightly more than half (54%) were male. The majority of the sample of preschoolers was African American (59%) or White (30%). Table 9.1 provides additional information on the demographic characteristics of the children in the study sample. There were no statistically detectable differences between the treatment and control groups on these child characteristics.

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 9.2. The average age of the primary caregiver was 31 years. About one-third (36%) were married, and 43 percent were never married. Approximately one-third (34%) of the primary caregivers reported having a high school diploma or GED; 13 percent had not finished high school; 38 percent had some college education; and 15 percent had a bachelor's degree or higher. More than half (63%) of the primary caregivers were employed full-time, 12 percent were employed part-time, and 23 percent were unemployed. There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

Teachers

There were 30 teachers who participated in the preschool-year intervention study. Most (97%) were female, and most were White (83%) or African American (13%). On average, the preschool teachers had 16 years of teaching experience, with an average of 9 years of experience teaching preschool. The majority of teachers had a bachelor's (53%) or graduate (27%) degree. An additional 13 percent had a high school diploma or GED. The majority of teachers reported having a current teaching license or certificate (80%). Some teachers also had a state-awarded preschool certificate (40%), or a Child Development Associate (CDA) credential (23%). Table 9.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 14 children. The child-staff ratio was on average 5.7 children to one teacher or program staff person in both locations.

Table 9.1. Child demographic characteristics for *Literacy Express* and *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*

Characteristics	Full sample n = 297	Curriculum comparison		
		Control n = 97	Treatment 1 ¹ n = 99	Treatment 2 ² n = 101
Age at baseline (years), mean	4.6	4.6	4.6	4.6
Gender (% male)	54.3	59.3	52.7	51.0
Race/ethnicity (%)				
White, non-Hispanic	29.6	23.5	40.2	25.5
African American, non-Hispanic	58.9	70.6	50.6	56.1
Hispanic	5.6	‡	5.7	8.2
Asian or Pacific Islander	‡	0.0	0.0	‡
Native American	‡	‡	‡	‡
Multiple/other	4.8	‡	‡	9.2
Child disability status (parent reported, %)	35.9	41.4	31.1	35.5

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹In Florida, Treatment 1 is *Literacy Express*.

²In Florida, Treatment 2 is *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 9.2. Primary caregiver demographic characteristics for *Literacy Express* and *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*

Characteristics	Full sample n = 268	Curriculum comparison		
		Control n = 86	Treatment 1 ¹ n = 90	Treatment 2 ² n = 92
Age at baseline (years), mean	31.2	30.0	31.6	31.8
Marital status (%)				
Married	36.2	24.4	41.1	42.4
Separated/Divorced	20.1	15.1	21.1	23.9
Widowed	‡	‡	‡	0.0
Never Married	42.5	59.3	35.6	33.7
Race/ethnicity (%)				
White, non-Hispanic	34.5	25.9	41.6	35.5
African American, non-Hispanic	58.4	71.8	50.6	53.8
Hispanic	4.9	‡	4.5	8.6
Asian or Pacific Islander	‡	0.0	‡	0.0
Native American	‡	‡	0.0	0.0
Multiple/other	‡	0.0	‡	‡
Educational level (%)				
Did not finish high school	13.2	23.5	9.1	7.5
High school diploma or GED	34.2	38.8	36.4	28.0
Some college	37.6	28.2	42.0	42.0
College graduate	15.0	9.4	12.5	22.6
Employment (%)				
Full-time	63.2	59.3	67.8	62.4
Part-time	11.5	10.5	13.3	10.8
Unemployed	23.4	27.9	17.8	24.7
Other	1.9	‡	‡	‡

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹In Florida, Treatment 1 is *Literacy Express*.

²In Florida, Treatment 2 is *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 9.3. Preschool teacher characteristics for *Literacy Express* and *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*

Characteristics	Full sample n = 30	Curriculum comparison		
		Control n = 9	Treatment 1 ¹ n = 10	Treatment 2 ² n = 11
Gender (% female)	97.0	89.0	100.0	100.0
Race/ethnicity (%)				
White, non-Hispanic	83.0	89.0	80.0	82.0
African American, non-Hispanic	13.0	0.0	‡	‡
Hispanic	‡	‡	0.0	0.0
Asian or Pacific Islander	0.0	0.0	0.0	0.0
Native American	0.0	0.0	0.0	0.0
Multiple/other	0.0	0.0	0.0	0.0
Educational level (%)				
High school diploma or GED	13.0	0.0	0.0	36.0
Associate's degree	‡	0.0	‡	‡
Bachelor's degree	53.0	78.0	50.0	36.0
Graduate degree	27.0	‡	40.0	‡
Current teaching license/certificate (%)	80.0	100.0	90.0	55.0
Child Development Associate (CDA) (%)	23.0	‡	‡	45.0
State-awarded preschool certificate (%)	40.0	‡	50.0	45.0
No credential (%)	0.0	0.0	0.0	0.0
Years of teaching experience, overall (mean)	15.9	17.6	15.4	15.1
Years of preschool teaching experience (mean)	9.3	10.7	10.4	7.1

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

¹In Florida, Treatment 1 is *Literacy Express*.

²In Florida, Treatment 2 is *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Random Assignment

The Florida-FSU research team recruited 17 schools to participate in the study. Schools were rank ordered according to their letter grade (A, B, C, or D) using Florida's school grading report⁵ for each school. It was important to consider school grade as a blocking variable, because the letter grades represent schools with percentages of students who are functioning at categorically different levels of academic achievement based on the Florida Department of Education's grading system. The sample of 17 schools included 11 grade A schools, one grade B school, three grade C schools, one grade D school, and one school for which a grade could not be determined. Within each letter grade ranking, the research team ranked each school by the average number of years of teaching experience that the teachers had. Once the list of 16 graded schools was rank-ordered, the schools were grouped into triplets, and within each triplet, the schools were randomly assigned (using the random function in Excel) to one of three conditions (*Literacy Express*, *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*, or control). The ungraded school and one

⁵ All schools in Florida receive a grade based on the following: (1) percentage of students meeting high standards of the Florida Comprehensive Assessment Test (FCAT)-achievement scores of Level 3 or above; (2) percentage of students making learning gains; and (3) adequate progress of the lowest 25 percent of the students in the school. Each school receives a certain number of points for each of three categories. The points are summed to create a total score. The total score is converted into a letter: grade A (410 points or more), grade B (380 to 409 points), grade C (320 to 379), grade D (280 to 319), and grade F (less than 250). For a grade of A, 95 percent were tested and at least 50 percent of the lowest readers must have made gains in the current school year. For a grade of B or C at least 50 percent of the lowest readers must have made gains in one or two consecutive years. Information reported here is based on the school grade categories in use during the 2003-04 school year. Source: <http://schoolgrades.fldoc.org>.

additional school that was a late entry to the project were randomly assigned separately to the *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K* curriculum condition and the control condition, respectively. Schools were randomly assigned to condition. The number of pre-kindergarten classrooms in each school ranged from one to three. Slightly more than half of the schools (9 of 17) had two pre-kindergarten classrooms per schools. At schools where there were two or more treatment group classrooms assigned to either *Literacy Express* or *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K* curriculum conditions, one of the two treatment group classrooms at those schools was then randomly assigned to a mentoring condition.⁶

Contamination

Because school was the unit of random assignment, all participating teachers within each school used the same curriculum, thus reducing the likelihood of contamination across conditions.

Control Condition

For all classrooms in the control condition, the school district was responsible for providing teachers with *High/Scope* curriculum training. The training provided to teachers in the control condition included a week-long summer institute conducted by *High/Scope* trainers prior to the start of the project, additional training sessions throughout the school year conducted by both *High/Scope* personnel and district personnel, and classroom visits by the *High/Scope* trainer. The evaluation, however, was not intended to be an evaluation of the *High/Scope* curriculum.

Data Collection

Mathematica Policy Research, Inc. (MPR) collected the child, parent, teacher, and school data for the Florida-FSU site for all three waves of data collection. The fall assessment data collection window for child assessments ranged from September 30, 2003 to November 17, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 42 days. The spring pre-kindergarten window was April 19, 2004 to June 15, 2004, and the kindergarten follow-up window was April 4, 2005 to June 6, 2005.

Attrition

Seventeen schools were randomly assigned to one of two treatment conditions or to the control condition. All 17 schools remained in the study throughout the pre-kindergarten year.

For the child assessment, the fall 2003 response rate was 95 percent; the spring 2004 pre-kindergarten response rate was 96 percent; and the kindergarten follow-up response rate was 80 percent.

Implementation

The Florida-FSU research team provided training and support to the treatment group teachers who implemented *Literacy Express* or *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*. Teachers and teachers' aides in the two treatment curriculum conditions were provided with all required materials and received direct training in the use of these curriculum materials. Curriculum training was provided to *Literacy Express* classroom teachers from July 28, 2003 to July 31, 2003, and for those implementing *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*, from July 30, 2003 to August 4, 2003. In each of the training sessions, the first 2 days were spent in a workshop setting and the other 2 were used for team planning. The workshop training session familiarized teachers and their aides with

⁶ Assignment to the mentoring conditions is a feature of the researcher's complementary research study. This report does not present findings based on the mentoring group assignment.

the new curriculum materials and provided hands-on experience in leading activities from the curricula. The workshop sessions were videotaped for those who were unable to attend to view at later times, so that all faculty and staff involved in the project would receive the same training. Make-up training using the videotapes was offered throughout the year as new staff joined the schools. Throughout the school year, teachers and teachers' aides received additional professional development opportunities in the use of the treatment curricula and related topics. All treatment group teachers attended a 2-hour professional development meeting specific to their assigned curriculum every other month. Attendance was documented at these meetings.

At sites (schools) where two or more treatment group classrooms were assigned to either *Literacy Express* or *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* curriculum conditions, one of the two treatment classrooms at each site was randomly assigned to a mentoring condition. Throughout the school year, teachers in the mentoring condition received visits in their classrooms from the project's mentor teacher. These visits lasted on average hours per week, for a monthly average of 8 to 10 hours per class. During the visits, the mentor teacher acted as a coach providing the teacher with the opportunity to engage in collegial conversation and receive extra support in the implementation of the curriculum. The mentor used a combination of techniques to provide support the teachers and teachers' aides. The techniques included demonstrations, feedback, and troubleshooting in the use of the appropriate curriculum. The mentor teacher also worked with individual students and groups of students who were not responding to the curricula, who were nonverbal, or who simply needed more intense intervention strategies to be successful in the curricula.

Site-specific curriculum fidelity observations were conducted in both treatment and control classrooms in February 2004, and April/May 2004. Observations in each classroom consisted of two observational rating systems (Early Language and Literacy Classroom Observation [ELLCO] and Center for Improving the Readiness of Children for Learning and Education [CIRCLE] teacher observation tool) as well as two specific fidelity measures for *Literacy Express* and *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*. The observational rating systems were completed following a 2.5 to 3-hour period of observation. The fidelity measures were completed every 20 to 30 minutes during the observation period. For 25 percent of the classrooms, a second observer completed these same fidelity and global ratings to provide an estimate of the reliability of the classroom measurement.

Implementation Fidelity Ratings

Literacy Express

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from "Not at All" (0) to "High" (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. *Literacy Express* was rated in the high Medium range (2.5) on the global implementation fidelity measure. The control group curriculum was rated at the Medium level (2.0).

DLM Early Childhood Express supplemented with Open Court Reading Pre-K

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from "Not at All" (0) to "High" (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* was rated in the high Medium range (2.3) on the global implementation fidelity measure. The control group curriculum was rated at the Medium level (2.0).

Impact Analysis Results

Because the Florida State University researchers evaluated two curricula, we present the results first for *Literacy Express* and then for *DLM Early Childhood Express* supplemented with *Open Court Reading Pre-K*. For each curriculum, we begin with the analyses of the child-level measures (i.e., mathematics, reading, phonological awareness, and language assessments) and then present the analyses of the classroom observation data.

Literacy Express—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-10a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-10a in appendix D. For all analyses of child-level measures, the following covariates were included (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 9.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three mathematics measures, we conclude that *Literacy Express* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically significant differences on the TERA and WJ Spelling test for the fall assessment. There was, however, a statistically significant difference (ESs = .44. $p < .05$) favoring the *Literacy Express* group on the WJ Letter Word Identification scale for the fall assessment. This difference could indicate the failure of randomization to achieve equivalent groups at the start of treatment or an early treatment effect. Additional analyses of these data are provided in appendix A.

There were no statistically detectable differences between groups on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three reading measures, we conclude that *Literacy Express* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups for the CTOPP spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Literacy Express* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

In the spring of the pre-kindergarten and kindergarten years, there were no statistically detectable differences between groups on either measure.

Based on the analyses of the two language measures, we conclude that *Literacy Express* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on the behavioral measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences between groups on any of the behavior measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) disability status as reported by the parent, (e) race/ethnicity, and (f) mother's education.

There was no statistically detectable difference between groups on the SSRS Social Skills and Problem Behaviors scales.

We obtained a statistically reliable impact on the Learning Behaviors Scale ($ES_s = -.38, p < .05$), such that children in the *Literacy Express* classrooms exhibited weaker learning behaviors relative to students in the control condition for the spring kindergarten assessment, but not the spring pre-kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Literacy Express* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition during pre-kindergarten or kindergarten assessments.

Literacy Express—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-10b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-10b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 9.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups for the fall observation.

There was a statistically detectable difference between the *Literacy Express* classrooms and the control classrooms on the ECERS-R in spring of pre-kindergarten ($ES_c = 1.29, p < .05$). Treatment group classrooms received higher global quality ratings relative to the control group classrooms.

Based on the analysis of the ECERS-R, we conclude that *Literacy Express* had a positive effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

There were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales for the spring pre-kindergarten observation.

Based on the analyses of the four Arnett scales, we conclude that *Literacy Express* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) for the spring pre-kindergarten assessment only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences on the Book Reading, Print and Letter Knowledge, Written Expression, Oral Language, or Math Concepts scales. There was a statistically significant difference on the Phonological Awareness ($ES_c = 1.26, p < .05$) scale, indicating that the *Literacy Express* teachers provided more instruction in phonological awareness relative to the instruction provided in the control classrooms.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Literacy Express* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Literacy Express* had a positive effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Literacy Express* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Literacy Express* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Literacy Express*

The findings for *Literacy Express* are summarized in table 9.4.

Table 9.4. Effect sizes for Literacy Express

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.05	-.02	—
CMA-A Mathematics Composite	-.02	-.21	—
Shape Composition ¹	-.01	-.14	—
Reading			
TERA	.17	-.11	—
WJ Letter Word Identification	.30	.08	—
WJ Spelling	.05	.06	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.14	†	.08
Language			
PPVT	.17	.16	—
TOLD	-.04	.10	—
Behavior			
SSRS Social Skills	-.06	†	-.37
SSRS Problem Behavior ²	-.31	†	.22
PLBS/LBS	.17	†	-.38*
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	1.29*	—	
Teacher-child interaction			
Arnett Detachment ³	-1.09	—	
Arnett Harshness ³	-.84	—	
Arnett Permissiveness ³	.51	—	
Arnett Positive Interactions	.56	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	.49	
TBRS Oral Language	†	.25	
TBRS Phonological Awareness	†	1.26*	
TBRS Print and Letter Knowledge	†	1.07	
TBRS Written Expression	†	-.03	
TBRS Math Concepts	†	-.12	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$ ¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

DLM Early Childhood Express supplemented with Open Court Reading Pre-K—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-11a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-11a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ES_s) are presented in table 9.5.

Mathematics assessments

We conducted repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, CMA-A Composite Score, and Shape Composition). There were no statistically detectable differences between groups on these measures for the fall assessment.

There were statistically reliable mean differences in scores on WJ Applied Problems for the spring pre-kindergarten assessment (ES_s = .36, $p < .01$)⁷ and the spring kindergarten assessment (ES_s = .48, $p < .001$) favoring children in the *DLM Early Childhood Express with Open Court Reading Pre-K* classrooms. There were no statistically detectable differences on the other two mathematics measures.

Based on the analyses for the three mathematics measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

We conducted repeated measures linear spline models to analyze the data from all three reading assessments. There were no statistically significant differences on the fall assessment for the TERA and the WJ Spelling. However, there was a statistically significant difference for the fall assessment on the WJ Letter Word Identification test (ES_s = .41, $p < .05$; follow-up analyses for this finding are included in appendix A).

There were statistically reliable mean differences on all three reading measures favoring students in the *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* classrooms in spring of pre-kindergarten: TERA (ES_s = .68, $p < .001$), WJ Letter Word (ES_s = .51, $p < .01$), and WJ Spelling (ES_s = .46, $p < .01$).

For the spring kindergarten assessment, statistically reliable differences were obtained on two of the three reading measures (TERA, ES_s = .76, $p < .01$; WJ Letter Word Identification, ES_s = .50, $p < .01$), indicating that the difference in spring of pre-kindergarten was sustained through spring of the following year. There was no statistically detectable difference in scores on the WJ Spelling.

Based on the analyses for the three reading measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* improved young children's early reading skills relative to the control condition.

Phonological awareness

The phonological awareness measures were the Pre-CTOPPP, Elision subtest, and the CTOPP, Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was a statistically reliable difference favoring the *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* condition in the spring of pre-kindergarten (Pre-CTOPPP, ES_s = .32, $p < .05$).

⁷ Significance indications (p -values) in the text refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was a statistically significant difference between groups on the CTOPP in the spring of kindergarten favoring the *DLM Early Childhood Express with Open Court Reading Pre-K* classrooms ($ES_s = .38, p < .05$).

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* improved phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were statistically reliable differences on the TOLD for the fall assessment ($ES_s = .38, p < .05$; follow-up analyses for this finding are included in appendix A).

In the spring of the pre-kindergarten year, there were statistically reliable mean differences in scores on both language measures (PPVT: $ES_s = .40, p < .05$; TOLD Grammatical Understanding: $ES_s = .40, p < .01$). These differences were sustained through spring of the following year (PPVT: $ES_s = .48, p < .01$; TOLD Grammatical Understanding subtest: $ES_s = .46, p < .01$).

Based on the analyses of the two language measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* improved children's language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) disability status as reported by the parent, (d) race/ethnicity, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment. For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

DLM Early Childhood Express supplemented with Open Court Reading Pre-K— Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-11b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-11b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 9.5.

Overall classroom environment

We conducted a repeated measures analysis on the ECERS-R. There was no statistically detectable difference between groups on the fall observation.

No statistically significant difference between groups was obtained for the spring pre-kindergarten assessment.

Based on the analysis of the ECERS-R, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

There were no statistically significant differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales on the spring of pre-kindergarten observation.

Based on the analyses of the four Arnett scales, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (TBRS Print and Letter Knowledge and Written Expression scales); (b) phonological awareness (TBRS Phonological Awareness scale); (c) language (TBRS Book Reading and Oral Language scales); and (d) early mathematics (TBRS Math Concepts scale) in the spring of the pre-kindergarten year only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences on the Book Reading, Print and Letter Knowledge, Written Expression, Oral Language, or Math Concepts scales. There was a statistically significant difference on the Phonological Awareness scale ($ES_c = 1.41$, $p < .05$), indicating that the *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* teachers provided more instruction in phonological awareness relative to teachers in the control classrooms.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had a positive effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

The findings for *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* are summarized in table 9.5.

Table 9.5. Effect sizes for *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

Measure	Student-level effect sizes (ES_s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.36**	.48***	—
CMA-A Mathematics Composite	.17	.13	—
Shape Composition ¹	.24	.09	—
Reading			
TERA	.68***	.76**	—
WJ Letter Word Identification	.51**	.50**	—
WJ Spelling	.46**	.22	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.32*	†	.38*
Language			
PPVT	.40*	.48**	—
TOLD	.40**	.46**	—
Behavior			
SSRS Social Skills	-.11	†	-.18
SSRS Problem Behavior ²	.11	†	.01
PLBS/LBS	-.16	†	-.13
	Classroom-level effect sizes (ES_c)		
Measure	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	.34	—	
Teacher-child interaction			
Arnett Detachment ³	-.06	—	
Arnett Harshness ³	-.70	—	
Arnett Permissiveness ³	.05	—	
Arnett Positive Interactions	.43	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	.01	
TBRS Oral Language	†	-.33	
TBRS Phonological Awareness	†	1.41*	
TBRS Print and Letter Knowledge	†	.91	
TBRS Written Expression	†	-.58	
TBRS Math Concepts	†	-.46	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

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

¹ Building Blocks, Shape Composition task

² Higher scores on this scale represent more negative child behaviors.

³ Lower scores on this scale represent a more positive classroom environment.

⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Chapter 10. Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: University of California, Berkeley/University at Buffalo, State University of New York (California/New York sites)

Curriculum

The University of California, Berkeley, and the University at Buffalo, State University of New York (California/New York) research team implemented the *Pre-K Mathematics* curriculum supplemented with the *DLM Early Childhood Express Math software* in preschool classrooms in California and New York.

The *Pre-K Mathematics* curriculum consisted of 29 small-group mathematics activities with concrete manipulatives for use by teachers and children in preschool classrooms as well as 19 home mathematics activities and materials for use by parents and preschool-age children in home settings. The teacher's manual provided a curriculum plan that linked small-group classroom activities to home activities.

Teachers conducted small-group mathematics activities twice per week with all pre-kindergarten children. Small-group activities are conducted with groups of 4 to 6 children for approximately 20 minutes per group. Teachers completed Assessment Record Sheets specifically tied to the mathematics activity during each small-group session. In addition to these structured activities, similar mathematics materials and activities were available to children in classroom mathematics centers for use during free play. Materials for home mathematics activities were sent home every 1 to 2 weeks and corresponded conceptually to the classroom mathematics activities.

The *DLM Early Childhood Express Math software* included 26 numerical, quantitative, geometric, and spatial activities. The *DLM Early Childhood Express Math software* is a component of Building Blocks, a research-based mathematics curriculum that addresses (a) geometric and spatial ideas and skills and (b) numeric and quantitative ideas and skills. Working with the *DLM Early Childhood Express Math software*, children use pattern blocks and tangrams to complete puzzles.

The software program provided individualized pre-kindergarten mathematics instructional activities for children to use approximately twice a week. Curriculum implementation was conducted over a 36-week period. Activities were scheduled such that children engaged in conceptually related small-group, home, and computer mathematics activities during the same week. Teachers were encouraged to present information from Assessment Record Sheets and to discuss children's mathematics learning during routine parent-teacher conferences.

Sample

The California/New York research team recruited five Head Start and public pre-kindergarten programs in California and two Head Start and public pre-kindergarten programs in New York. A total of 40 teachers/classrooms (20 in each state) were recruited from these Head Start and public pre-kindergarten programs to participate in the study. Twenty-six (12 in California and 14 in New York) of the 40 classrooms were full-day pre-kindergarten programs. Consent letters were sent home to the parents of all eligible children in each classroom. Teachers and other classroom staff assisted with the recruitment of families. A sample of 316 children (159 treatment, 157 control) and parents were recruited for participation in the study. Data were collected on a total of 314 children and 263 parents at the time of the fall assessment.

The kindergarten sample included 309 children. Data were collected on 283 children and 246 parents at the time of the spring kindergarten assessment.

Children and Families

The children were 4.3 years of age at the time of the fall assessment data collection and almost half (48%) were male. The sample included African American (45%), Hispanic (23%), and White (18%) preschoolers. The racial/ethnic composition of the sample of children varied based on the geographic location of the sample. The California sample was primarily African American (48%) or Hispanic (35%). A larger percentage of White children (36%) were represented in the New York sample. Table 10.1 provides additional information on the demographic characteristics of the children in the California and New York study samples. At baseline, there were more boys in the control group classrooms relative to those assigned to the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* classrooms (52% vs. 43%, $p = .05$).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 10.2. The average age of the primary caregiver was 32 years. Almost half (43%) of the primary caregivers were never married; 40 percent were married at the time of the fall assessment data collection. Half of the primary caregivers reported having had some college education (40%) or a college degree (11%); 27 percent had a high school diploma or GED; and 23 percent had not finished high school. A large percentage (40%) of the primary caregivers were not working at the time of the fall data collection. Some were employed full-time (37%) or part-time (20%). There were no statistically detectable differences between the treatment and control groups on the primary caregiver characteristics.

Teachers

Forty teachers participated in the preschool year intervention study; all were female. The racial/ethnic composition of the sample included White (38%), African American (33%), Hispanic (13%), and Asian (10%) teachers. On average, the preschool teachers had 19 years of teaching experience, with an average of 12 years of experience teaching preschool. The majority of teachers had a bachelor's (33%) or graduate (40%) degree. An additional 18 percent had an associate's degree, and 10 percent had a high school diploma or GED. The majority of teachers reported having a current teaching license/certificate (78%); state-awarded preschool certificate (68%); or a Child Development Associate (CDA) credential (33%). Table 10.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 22.4 children in California, and 14.4 children in New York. The child-staff ratio was on average 7 to 1 in California, and 6.7 to 1 in New York.

Table 10.1. Child demographic characteristics for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

Characteristics	Full sample n = 316	Curriculum comparison	
		Control n = 157	Treatment n = 159
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: California and New York</i>			
Age at baseline (years), mean	4.3	44.4	4.4
Gender (% male)	47.5	52.2*	42.7
Race/ethnicity (%)			
White, non-Hispanic	17.9	13.8	21.9
African American, non-Hispanic	44.7	49.0	40.4
Hispanic	23.0	22.8	23.3
Asian or Pacific Islander	3.1	‡	5.5
Native American	‡	‡	‡
Multiple/other	10.3	13.1	7.5
Child disability status (parent reported, %)	10.3	9.6	11.0
Characteristics	Full sample n = 160	Curriculum comparison	
		Control n = 80	Treatment n = 80
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: California</i>			
Age at baseline (years), mean	4.5	4.4	4.5
Gender (% male)	46.9	57.5	36.3
Race/ethnicity (%)			
White, non-Hispanic	‡	‡	‡
African American, non-Hispanic	47.7	41.0	54.7
Hispanic	34.6	38.5	30.7
Asian or Pacific Islander	5.9	‡	10.7
Native American	0.0	0.0	0.0
Multiple/other	10.5	17.9	‡
Child disability status (parent reported, %)	6.7	7.9	5.4
Characteristics	Full sample n = 154	Curriculum comparison	
		Control n = 77	Treatment n = 77
<i>Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: New York</i>			
Age at baseline (years), mean	4.3	4.3	4.3
Gender (% male)	48.1	46.8	49.4
Race/ethnicity (%)			
White, non-Hispanic	36.2	28.4	43.7
African American, non-Hispanic	41.3	58.2	25.4
Hispanic	10.1	‡	15.5
Asian or Pacific Islander	0.0	0.0	0.0
Native American	‡	‡	‡
Multiple/other	10.1	7.5	12.7
Child disability status (parent reported, %)	15.3	12.2	17.7

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

* $p < .05$

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 10.2. Primary caregiver demographic characteristics for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

Characteristics	Full sample n = 261	Curriculum comparison	
		Control n = 125	Treatment n = 136
Age at baseline (years), mean	32.5	31.9	33.0
Marital status (%)			
Married	39.6	43.2	36.3
Separated/Divorced	15.0	12.8	17.0
Widowed	2.3	3.2	‡
Never Married	43.1	40.8	45.2
Race/ethnicity (%)			
White, non-Hispanic	21.7	18.5	24.6
African American, non-Hispanic	42.6	46.0	39.6
Hispanic	22.5	25.0	20.1
Asian or Pacific Islander	4.7	‡	8.2
Native American	‡	‡	‡
Multiple/other	7.4	8.9	6.0
Educational level (%)			
Did not finish high school	22.7	27.4	18.4
High school diploma or GED	26.9	30.6	23.5
Some college	39.6	33.1	45.6
College graduate	10.8	8.9	12.5
Employment (%)			
Full-time	37.2	32.8	41.2
Part-time	19.9	17.6	22.1
Unemployed	40.2	46.4	34.6
Other	2.7	3.2	‡

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 10.3. Preschool teacher characteristics for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

Characteristics	Full sample n = 40	Curriculum comparison	
		Control n = 20	Treatment n = 20
Gender (% female)	100.0	100.0	100.0
Race/ethnicity (%)			
White, non-Hispanic	38.0	40.0	37.0
African American, non-Hispanic	33.0	40.0	26.0
Hispanic	13.0	‡	‡
Asian or Pacific Islander	10.0	0.0	21.0
Native American	0.0	0.0	0.0
Multiple/other	‡	‡	0.0
Educational level (%)			
High school diploma or GED	10.0	‡	‡
Associate's degree	18.0	20.0	‡
Bachelor's degree	33.0	40.0	25.0
Graduate degree	40.0	35.0	45.0
Current teaching license/certificate (%)	78.0	80.0	75.0
Child Development Associate (CDA) (%)	33.0	35.0	32.0
State-awarded preschool certificate (%)	68.0	74.0	12.0
No credential (%)	‡	‡	0.0
Years of teaching experience, overall (mean)	19.0	19.5	18.5
Years of preschool teaching experience (mean)	12.4	13.4	11.4

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Random Assignment

A total of 40 Head Start and public preschools were randomly assigned in the fall of the pilot study year by the research team, using block randomization to either the treatment condition (*Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*) or the control condition. Blocks were formed at the program level (five programs in California and two in New York), with teachers from Head Start and state-funded programs balanced by curriculum assignment in each site. In California, random assignment of classrooms to intervention and control conditions was done publicly in the presence of project staff and program staff for each of the five programs. The names of teachers who expressed a willingness to participate in the study were placed in a container and randomly drawn. The classroom of the first teacher whose name was drawn was assigned to the treatment condition. The classroom of the second teacher whose name was drawn was assigned to the control condition. This random assignment process continued until the designated number of classrooms had been assigned to each condition. To conduct a substudy of children from low-income Asian American families, two classrooms with large enrollments of Chinese American children were paired, such that when the teacher for one classroom was randomly assigned to a condition, the other classroom was automatically assigned to the other condition. Furthermore, two classrooms in which the language of instruction was Spanish were paired, such that when one classroom was randomly assigned to one condition (e.g., treatment) the other classroom was automatically assigned to the other condition (control). In New York, the school district and Head Start program administrators provided the research team with the names of teachers who were willing to participate in the study. Assignment to treatment and control conditions was then done publicly. The names were then randomly drawn, with the stipulation that there could not be a treatment and control teacher in the same building. A subsample of eight focal children was randomly selected in each classroom from the total number of consented children, balanced for age and gender. A total of 40 classrooms and 315 were recruited to participate in the study.

The same assignments were maintained for the second year (2003-04) of implementation as had been used during the pilot-study year (2002-03). In California, the research team was able to retain 8 of the 10 treatment classrooms and nine of the 10 control classrooms from the pilot year. They were able to retain 8 of the 10 treatment group teachers and 8 of the 10 control group teachers. Three preschool classrooms (two treatment and one control) were closed or converted by the program to serve age groups other than pre-kindergarten. Three replacement classrooms and their head teachers were added by randomly selecting them from the list of volunteers provided by the participating programs. In another control classroom, the teacher left after the pilot year and the program assigned a new teacher to the classroom. This classroom was retained in the control condition.

In New York, the research team was able to retain 8 of the 10 pilot-year treatment group teachers. They were able to retain 9 of the 10 control group teachers. Three replacement teachers (two in the treatment condition and one in the control condition) were added to the study sample. These three teachers were assigned to the existing treatment and control group classrooms (based on the initial intention that the classroom was the unit of random assignment). After the beginning of the school year, one of the treatment schools closed its pre-kindergarten program and the teacher, one of the replacements teachers, was reassigned. The district pre-kindergarten administrator was contacted and provided the research team with the original list of volunteers. A replacement teacher was randomly selected from this list.

Contamination

The California/New York researchers randomly assigned classrooms/teachers in each location to the intervention or control condition. To minimize the likelihood of contamination, intervention and control classrooms were located in different buildings. Furthermore, programs were asked to ensure that intervention and control teachers did not substitute in classrooms assigned to a condition different from their own. Finally, classroom observations using the Early Mathematics Classroom Observation instrument (described below), as well as periodic unannounced classroom visits in treatment and control classrooms by project staff did not reveal any evidence of contamination.

Control Condition

A number of curricula were represented in the control classrooms. Prevailing curricula included *Creative Curriculum*, *High Scope*, *Montessori*, specialized literacy curricula, and local school district and teacher-developed curricula. In New York, control group teachers in the public pre-kindergarten classrooms used the *BPS Benchmarks*, a curriculum that was developed by the local school district. Head Start classroom teachers in New York used a version of the *Creative Curriculum*.

Data Collection

RTI International (RTI) collected the child, teacher, and school data for the California and New York sites for all three waves of data collection. The California/New York research team was responsible for conducting the parent interviews in the preschool year, except for the few Chinese-speaking parents who were interviewed by a trained member of the grantee staff who spoke Chinese. In the kindergarten follow-up year, RTI staff completed the parent interviews. The fall assessment data collection window for child assessments ranged from September 22, 2003 to November 7, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 19 days in California and 14 days in New York. The spring pre-kindergarten window was April 7, 2004 to June 10, 2004, and the kindergarten follow-up window was May 2, 2005 to June 14, 2005.

The California/New York researchers supplemented the assessment of mathematics practices by intervention and control teachers by administering the Early Mathematics Classroom Observation (EMCO) (Klein and Starkey 2000). The EMCO measured the *amount* of classroom support for mathematical development by recording the number of children who participated in a mathematics activity and the duration of their

participation, thus yielding a measure of minutes-of-math support per child for a given classroom. This instrument also measured the *nature* of classroom support for mathematical development: (1) the conceptual domain supported, such as number and arithmetic or spatial and geometric knowledge; (2) the specific concepts and skill supported, such as cardinal number and counting sets of 1-10 objects; and (3) the general type of mathematics activity presented by the teacher (focal mathematics, which has a primarily mathematical goal, or embedded mathematics, such as a cooking activity, which includes some mathematics, but has a primarily nonmathematical goal).

The California/New York researchers also administered the Child Math Assessment (CMA) at pretest and posttest. The CMA is a comprehensive assessment of early mathematical knowledge. The assessment was comprised of 17 tasks, using concrete objects and encompassing a range of problem difficulty appropriate for pre-kindergarten children. This instrument assesses mathematical knowledge within several distinct areas, including number, arithmetic, space and geometry, measurement, and pattern knowledge.

Attrition

Forty classrooms were randomly assigned to treatment or control condition. All 40 classrooms remained in the study from the beginning of the pre-kindergarten year through the spring of the pre-kindergarten year.

For the child assessment, the fall assessment response rate was 99 percent, the spring 2004 pre-kindergarten response rate was 94 percent, and the kindergarten follow-up response rate was 92 percent.

Implementation

Most (16 of 20) teachers in the treatment condition were in their second year of implementation of the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* at the time of the evaluation. The teacher sample included teachers who participated in the pilot year of the study (2002-03), and new teachers who started in 2003-04. During the pilot year, treatment group teachers in California participated in 4-day training workshops in late summer (September 10-13) and winter (February 3-6). In New York, treatment group teachers participated in 4-day training workshops in fall (September 26-27; October 3-4) and in winter (February 10-11 and 20-21). Ongoing on-site training was provided by project staff approximately twice per month, for an average of 17 on-site training sessions per teacher in California and 12.5 training sessions per teacher in New York. During the second year of the implementation, treatment group teachers attended a refresher workshop for 2 days in late summer. Throughout the preschool year, project staff observed and rated the implementation fidelity of small-group activities in each intervention classroom 1-2 times per month. Feedback was given to treatment group teachers at the end of those observation sessions. Staff members also observed teachers and children while they were using the computer-based mathematics activities, examined computer records of children's use of these activities, and provided feedback and training to teachers as needed.

The California/New York research team collected fidelity of implementation data, using the Fidelity of Implementation Record Sheet (Klein and Starkey 2002), as part of their formative evaluation of the mathematics curriculum. They collected data on the fidelity of implementation of small-group activities, computer activities, and home activities. Implementation fidelity data were collected in fall and spring of the pre-kindergarten year. The research team also administered the Early Mathematics Classroom Observation (Klein and Starkey 2000) to collect data on the amount of teacher-participant mathematics support per child that was provided in treatment and control classrooms, whether mathematics content was focal or embedded on other types of activities, and the conceptual breadth of mathematics support provided by teachers. For the home activities measure, parents were asked to report on how often mathematics activities were sent home, how often they used the activities, whether they liked the activities, whether the activities helped their children learn mathematics, and whether the activities gave them ideas about how to help their children learn mathematics.

On the site-specific fidelity measure, the overall fidelity scores for the small-group activities and computer mathematics activities were calculated by averaging scores from one fall and one spring fidelity observation. Small-group mathematics fidelity ranged from Moderate to High across classrooms, and computer mathematics fidelity ranged from low Moderate to High. Overall levels of fidelity were similar across the California (average = .87, Head Start classrooms; .92 state pre-kindergarten classrooms) and New York (average = .78, Head Start classrooms; .84 state pre-kindergarten classrooms) research sites.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Treatment implementation was rated between High and Medium in California (2.65) and New York (2.25) on the global fidelity measure. Researchers were also asked to provide a global rating for the control group classrooms. The fidelity of implementation ratings for the various control group curricula was at the Medium level (2.0) at both sites.

Impact Analysis Results

We begin with the analyses of the child-level measures (i.e., mathematics, reading, phonological awareness, and language assessments) and then present the analyses of the classroom observation data. Our discussion of the results focuses on the combined analysis for the California and New York sites.

Pre-K Mathematics supplemented with DLM Early Childhood Express Math software—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-12a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-12a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child’s age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother’s education. The student-level effect sizes (ES_s) are presented in table 10.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically significant differences for the fall assessment for the WJ Applied Problems and the CMA-A Composite Score. However, there was a statistically significant difference for the fall assessment on the Shape Composition task (ES_s = .25, $p < .05$; follow-up analyses are included in appendix A).

There was no statistically detectable difference for the spring pre-kindergarten or kindergarten assessments on the WJ Applied Problems.

In spring of the pre-kindergarten year, there was a statistically reliable mean difference in scores on the CMA-A Composite Score (ES_s = .44, $p < .01$) favoring the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* group. No difference was found for the CMA-A Composite Score for the spring kindergarten assessment.

In spring of the pre-kindergarten year, there was a statistically reliable mean difference in scores on Shape Composition (ES_s = .96, $p < .001$) favoring the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* group. The advantage of the treatment group was maintained through spring of the kindergarten year (ES_s = .41, $p < .001$).

Based on the analyses for the three mathematics measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* had a positive effect on children's early mathematics skills at the end of pre-kindergarten relative to the control condition. *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on children's mathematics skills at the end of kindergarten.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on any of the three reading measures at the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the three reading measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT], Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences at the fall assessment.

In the spring pre-kindergarten and kindergarten years, there were no statistically detectable differences between groups on either measure.

Based on the analyses of the two language measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were child's age, gender, disability status as reported by the parent, race/ethnicity, and mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Pre-K Mathematics supplemented with DLM Early Childhood Express Math software—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-12b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-12b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 10.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups on the fall observation.

No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

There were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales in spring of the pre-kindergarten year.

Based on the analyses of the four Arnett scales, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales); (b) phonological awareness (TBRS Phonological Awareness scale); (c) language (TBRS Book Reading and Oral Language scale); and (d) early mathematics (TBRS Math Concepts scale) in spring of pre-kindergarten only. To analyze these data, ANCOVAs were conducted. The covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The effect sizes are presented in table 10.4.

There were no statistically detectable differences on the TBRS Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, or Math Concepts scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

The findings for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* are summarized in table 10.4.

Table 10.4. Effect sizes for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.22	.13	—
CMA-A Mathematics Composite	.44**	.13	—
Shape Composition ¹	.96***	.41***	—
Reading			
TERA	.13	.31	—
WJ Identification	-.01	.22	—
WJ Spelling	.20	.03	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.04	†	-.11
Language			
PPVT	.17	.11	—
TOLD	.17	.08	—
Behavior			
SSRS Social Skills	.22	†	.06
SSRS Problem Behavior ²	-.09	†	-.01
PLBS/LBS	.09	†	.01
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	.05	—	
Teacher-child interaction			
Arnett Detachment ³	-.37	—	
Arnett Harshness ³	.18	—	
Arnett Permissiveness ³	-.45	—	
Arnett Positive Interactions	.16	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	.07	
TBRS Oral Language	†	.19	
TBRS Phonological Awareness	†	.38	
TBRS Print and Letter Knowledge	†	.07	
TBRS Written Expression	†	-.12	
TBRS Math Concepts	†	.57	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

** $p < .01$; *** $p < .001$

¹ Building Blocks, Shape Composition task

² Higher scores on this scale represent more negative child behaviors.

³ Lower scores on this scale represent a more positive classroom environment.

⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Chapter 11. *Project Approach*: Purdue University and University of Wisconsin-Milwaukee (Wisconsin site)

Curriculum

The Purdue University and University of Wisconsin-Milwaukee (Purdue/Wisconsin) research team implemented the *Project Approach* curriculum. *Project Approach* is a set of teaching strategies that enables teachers to guide children through in-depth investigations of real world topics. The curriculum is designed to use children's interests as the starting point for organizing and developing classroom learning activities. There are three curriculum components that address children's learning needs: spontaneous play, systematic instruction, and project work.

A *project* is defined as an in-depth study of a real world topic that is worthy of children's attention and effort. Projects can be incorporated into an existing classroom instructional program and can extend over several days or weeks. The structural features of the *Project Approach* include discussion, fieldwork, representation, investigation, and display. During the preliminary planning stage, the teacher selects the topic of study (based primarily on classroom learning goals, children's interests, and the availability of local resources). The teacher then brainstorms her own experience, knowledge, and ideas and represents them in a topic web. This topic web is revised throughout the project and used for recording the progress of the project. In *Project Approach* classrooms, the daily schedule is to be structured so that children and teachers spend at least 45 to 60 minutes engaged in investigation and discovery, typically in small groups.

Sample

The Purdue/Wisconsin research team recruited public pre-kindergarten classrooms for participation in the study. The research team recruited 13 teachers from 12 different schools. The recruitment of parents and children began at the start of the preschool year and continued through the first 6 weeks of school. Teachers assisted with the recruitment of families. Parents were offered an incentive for completing the parent interviews. A sample of 204 children (114 treatment, 90 control) and parents were recruited for participation in the study. Data were collected on 204 children and 176 parents at the time of the fall baseline data collection.

In the follow-up year of the evaluation, the sample of schools went from 12 in pre-kindergarten to more than 37 in kindergarten. The sample of classrooms went from 13 preschool to 58 kindergarten classrooms. The kindergarten sample included 156 children and 153 parents from the original sample of participants. Data were collected on 150 children and 122 parents.

Children and Families

The children were 4.6 years of age at the time of baseline data collection and slightly more than half (53%) were male. The racial/ethnic composition of the sample was diverse: African American (40%), White (28%), and Hispanic (17%). Table 11.1 provides additional information on the demographic characteristics of the children in the study sample. At baseline, the treatment group had a higher percentage of African American children relative to the control group (52% vs. 24%, $p < .01$).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 11.2. The average age of the primary caregiver was 31.7 years. Half (50%) the primary caregivers were married and 38 percent were never married. Half (51%) reported having had some college or a college degree; 32 percent had a high school diploma or GED; and 17 percent had not finished

Table 11.1. Child demographic characteristics for *Project Approach*

Characteristics	Full sample n = 204	Curriculum comparison	
		Control n = 90	Treatment n = 114
Age at baseline (years), mean	4.6	4.6	4.6
Gender (% male)	52.9	56.7	50.0
Race/ethnicity (%)			
White, non-Hispanic	28.2	36.6	21.2
African American, non-Hispanic	39.8	24.4	52.5**
Hispanic	17.1	20.7	14.1
Asian or Pacific Islander	‡	0.0	‡
Native American	‡	0.0	‡
Multiple/other	13.3	18.3	9.1
Child disability status (parent reported, %)	17.7	16.5	18.8

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

** $p < .01$

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 11.2. Primary caregiver demographic characteristics for *Project Approach*

Characteristics	Full sample n = 176	Curriculum comparison	
		Control n = 80	Treatment n = 96
Age at baseline (years), mean	31.7	32.0	31.4
Marital status (%)			
Married	50.0	63.8*	38.5
Separated/Divorced	11.4	11.3	11.5
Widowed	‡	0.0	‡
Never Married	38.1	25.0	49.0
Race/ethnicity (%)			
White, non-Hispanic	40.0	51.9	30.2
African American, non-Hispanic	41.7	25.3	55.2
Hispanic	9.1	12.7	6.3
Asian or Pacific Islander	‡	0.0	‡
Native American	2.3	‡	‡
Multiple/other	6.3	7.6	5.2
Educational level (%)			
Did not finish high school	17.2	14.1	19.8
High school diploma or GED	32.2	24.4	38.5
Some college	29.9	35.9	25.0
College graduate	20.7	25.6	16.7
Employment (%)			
Full-time	52.3	51.3	53.1
Part-time	18.8	25.0	13.5
Unemployed	27.8	21.3	33.3
Other	‡	‡	0.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

* $p < .05$

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

high school. More than half (52%) of the primary caregivers were employed full-time, 19 percent were employed part-time, and 27.8 percent were unemployed. At baseline, a higher percentage of parents were married in the control curriculum relative to those assigned to *Project Approach* (64% vs. 39%, $p < .05$).

Teachers

There were 12 teachers who participated in the preschool year intervention study. All of the preschool teachers were female, and all were White. On average, the preschool teachers had 11 years of teaching experience, with an average of 8 years of experience teaching preschool. All of the teachers had a bachelor's (54%) or graduate (46%) degree, and all reported having a current teaching license/certificate. Table 11.3 provides additional information on the characteristics of the preschool sample of teachers. At baseline, a higher percentage of teachers in the control had more years of teaching experience (17 vs. 6, $p < .01$), and years of preschool teaching experience (12 vs. 5, $p < .05$) relative to those assigned to *Project Approach*.

Table 11.3. Preschool teacher characteristics for *Project Approach*

Characteristics	Full sample n = 13	Curriculum comparison	
		Control n = 6	Treatment n = 7
Gender (% female)	100.0	100.0	100.0
Race/ethnicity (%)			
White, non-Hispanic	100.0	100.0	100.0
African American, non-Hispanic	0.0	0.0	0.0
Hispanic	0.0	0.0	0.0
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	0.0	0.0	0.0
Educational level (%)			
High school diploma or GED	0.0	0.0	0.0
Associate's degree	0.0	0.0	0.0
Bachelor's degree	54.0	‡	71.0
Graduate degree	46.0	67.0	‡
Current teaching license/certificate (%)	100.0	100.0	100.0
Child Development Associate (CDA) (%)	‡	0.0	‡
State-awarded preschool certificate (%)	36.0	‡	‡
No credential (%)	0.0	0.0	0.0
Years of teaching experience, overall (mean)	11.5	17.3**	6.4
Years of preschool teaching experience (mean)	8.4	12.3*	5.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

* $p < .05$; ** $p < .01$

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Programs/Classrooms

The average preschool class size was 24.2 children. The child-staff ratio was on average 15.2 children to one teacher or program staff person.

Random Assignment

Randomization was done during the pilot year of curriculum implementation. Twelve of 57 eligible teachers agreed to participate in the study. They were randomly assigned to the treatment and control conditions after stratifying for racial/ethnic composition of families served by the schools. However, because of teacher attrition between the pilot year and the second year of implementation, the pilot-year randomization was not maintained in year 2. In the second year of implementation, the district administrator provided the research team with a list of eligible schools from which to recruit a study sample for the second year of the study. Two pilot-year control group teachers and 11 newly recruited teachers were randomly assigned to 7 treatment and 6 control classrooms. The Purdue/Wisconsin research team randomly assigned 13 teachers to the experimental conditions (7 treatment and 6 control classrooms). The names of the teachers were placed in a container and randomly drawn and assigned to either the treatment or control group. In all schools but one, there was only one preschool classroom. In one school with two classrooms, both classrooms/teachers were assigned to the same condition (the treatment group) to avoid contamination. In all other schools, only one teacher/classroom was assigned to either the treatment or control condition. A total of 13 classrooms and 204 children took part in the study.

Contamination

Because all of the classrooms in each school were assigned to either the treatment or the control group, there was little risk of contamination across the treatment and control conditions.

Control Condition

The school district provided all preschool classrooms with the *Doors to Discovery* and *Growing with Mathematics* curriculum materials but these curricula were not used consistently across all of the classrooms. In the control classrooms, teachers reported implementing their own teacher-developed, nonspecific curricula when the research team asked them to report on the curriculum in use.

Data Collection

RTI International (RTI) collected the child, teacher, and school data for the Wisconsin site for all three waves of data collection. The Purdue/Wisconsin research team was responsible for conducting the parent interviews in the preschool year. In the kindergarten follow-up year, RTI staff completed the parent interviews. The fall assessment data collection window for child assessments ranged from September 15, 2003 to October 31, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 13 days. The spring pre-kindergarten window was April 11, 2004 to June 10, 2004, and the kindergarten follow-up window was April 2, 2005 to June 6, 2005.

Attrition

Thirteen classrooms were randomly assigned to treatment or control condition. All 13 classrooms remained in the study from the beginning of the pre-kindergarten year through the spring of the pre-kindergarten year.

For the child assessment, the baseline (fall, 2003) response rate was 100 percent; the spring 2004 response rate was 94 percent; and the kindergarten follow-up response rate was 96 percent.

Implementation

The research team provided training and support to the treatment group teachers to implement the *Project Approach* curriculum. On average, each treatment group teacher received 48 hours of training and individualized support during the 2003-04 preschool year (October 2003 through May 2004). The training and support activities included 18 hours of introductory training; 12 hours of follow-up training; and an

average of 12 hours of individual consultation time with the curriculum mentor during the mentor's regular visits to intervention classrooms.

The introductory training was held for 3 days (6 hours per day) at the beginning of the preschool year. The training workshop was conducted by an expert on the *Project Approach* and a co-author of *Young Investigators: The Project Approach in the Early Years*, the primary source of information on the *Project Approach* (Helm and Katz 2001). The content of the introductory training focused on benefits of the *Project Approach*; distinctions between projects and themes; criteria for selecting a good project topic; a detailed examination of the three phases of projects; and the use of webbing to link curriculum goals to project work. On the final day of the introductory training, participants visited a classroom in a local school (ineligible for the study) with a curriculum that included projects.

The follow-up training was held for 2 days in January 2004 (6 hours per day). The training began with a visit to study participant who demonstrated a high level of *Project Approach* implementation. The other treatment teachers toured her classroom, heard a presentation on the projects done to date in the classroom, and participated in a focused review of how various project activities were connected to curriculum goals for 4-year-old children. One goal of the visit was to help teachers strengthen integration of the *Project Approach* with other curriculum activities, including reading/writing, mathematics, and science. The follow-up training also included a presentation and critique of recent and/or on-going projects by each teacher, and a problem-solving discussion focused on challenges and barriers encountered by each teacher in implementing projects. In the final session of the follow-up training, each teacher generated an anticipatory planning web for the next project in her classroom, and received feedback and guidance from the trainer and other teachers.

In addition to group training sessions, the curriculum mentor conducted an average of 20.7 curriculum-related visits to each treatment classroom from October to May of the preschool year. On average, each mentoring visit was 2.8 hours in length, with 21 minutes of this time devoted to individualized consultation with the teacher about curriculum implementation. The rest of the mentor's time for each visit was devoted to classroom observation. The content of the mentoring visits focused primarily on: clarifications and reminders regarding components of the *Project Approach*; suggestions and feedback regarding planning and/or implementing project work (e.g., suggestions for experts and field visits); and provisions of resources to support project work (e.g., pizza recipes for children to use in a pizza project). The mentor completed a form for each visit that documented the observation of project-related displays and activities; specific feedback provided to the teacher based on the observation; suggestions offered to the teacher; resources provided by the mentor to the classroom; and specific actions the teacher agreed to take.

Site-specific curriculum fidelity data were collected three times in each of the seven classrooms implementing the *Project Approach*. The observations were conducted at three time points across the school year (November 2003 to January 2004; February to March 2004; and April to May 2004). Curriculum fidelity was measured with an observation and interview protocol. An experienced early childhood educator was trained by the project Principal Investigator to conduct the observations. The measure included items to address 12 main components: number of different types of displays; engaging, accessible displays; supports for project work; frequency of project work; level of engagement in project work; level of engagement in project-related work during free play/work time; level of child interest in project topic; use of experts and field visits; number and frequency of activities and materials; number and frequency of mathematics-related experiences; extent of opportunities for parent involvement; and teacher planning and documentation. The research team did not conduct curriculum fidelity observations in the control classrooms because of the wide variability in the use of curriculum materials in these classrooms. The school district provided all preschool classrooms with the *Doors to Discovery* and *Growing with Mathematics* curriculum materials, but these curricula were not used consistently across all of the classrooms. The research team collected observational and interview data on control classroom teachers' use of themes as part of their complementary research.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. *Project Approach* was rated Medium (1.86) on the global implementation fidelity measure. The control group curriculum was also rated at the Medium level (2.00).

Impact Analysis

We begin with the analyses of the child-level measures (i.e., mathematics, reading, phonological awareness, language, and behavioral assessments) followed by classroom measures.

Project Approach—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-13a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-13a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child’s age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother’s education. The student-level effect sizes (ESs) are presented in table 11.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences between groups on the spring pre-kindergarten or kindergarten assessments on any of the mathematics assessments.

Based on the analyses for the three mathematics measures, we conclude that *Project Approach* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three reading measures, we conclude that *Project Approach* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d) race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Project Approach* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall assessment.

There were no statistically detectable differences on either of these measures for the spring pre-kindergarten or spring kindergarten assessment.

Based on the analyses of the two language measures, we conclude that *Project Approach* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

For the spring kindergarten assessments, teachers rated children who had received *Project Approach* in pre-kindergarten as exhibiting more Problem Behaviors ($ES_s = .49, p < .05$), having weaker Social Skills ($ES_s = -.44, p < .05$), and fewer learning behaviors ($ES_s = -.42, p < .05$), relative to children from the pre-kindergarten control classrooms.

Based on the analyses of the behavioral measures, we conclude that *Project Approach* had a negative effect on children's social and learning behaviors in kindergarten, but not during pre-kindergarten.

Project Approach—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-13b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-13b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 11.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups on the fall observation.

No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Project Approach* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

There were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales in spring of the pre-kindergarten year.

Based on the analyses of the four Arnett scales, we conclude that *Project Approach* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales), (b) phonological awareness (TBRS Phonological Awareness scale), (c) language (TBRS Book Reading and Oral Language scales), and (d) early mathematics (TBRS Math Concepts scale) for the spring pre-kindergarten assessment only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences between groups on any of the TBRS scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Project Approach* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Project Approach* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Project Approach* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Project Approach* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Project Approach*

The findings for *Project Approach* are summarized in table 11.4.

Table 11.4. Effect sizes for *Project Approach*

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.07	.27	—
CMA-A Mathematics Composite	.18	.22	—
Shape Composition ¹	.27	.24	—
Reading			
TERA	.14	.29	—
WJ Letter Word Identification	.42	.03	—
WJ Spelling	.27	.14	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.05	†	-.17
Language			
PPVT	.16	.10	—
TOLD	.15	.32	—
Behavior			
SSRS Social Skills	.04	†	-.44*
SSRS Problem Behavior ²	.50	†	.49*
PLBS/LBS	-.31	†	-.42*
	Classroom-level effect sizes (ES _c)		
Measure	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	-.19	—	
Teacher-child interaction			
Arnett Detachment ³	.57	—	
Arnett Harshness ³	.86	—	
Arnett Permissiveness ³	-.43	—	
Arnett Positive Interactions	-.99	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	-.76	
TBRS Oral Language	†	-.42	
TBRS Phonological Awareness	†	-1.19	
TBRS Print and Letter Knowledge	†	.34	
TBRS Written Expression	†	.62	
TBRS Math Concepts	†	-.64	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$ ¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Page intentionally left blank.

Chapter 12. *Project Construct*: University of Missouri-Columbia (Missouri site)

Curriculum

The University of Missouri (Missouri) research team evaluated the *Project Construct* curriculum. *Project Construct* was developed under the direction of the Missouri Department of Elementary and Secondary Education in 1986 to fulfill the need for a curriculum and assessment framework that supports children’s learning. *Project Construct* is derived from constructivism—the theoretical view that learners construct knowledge through interactions with the physical and social environments. The preschool curriculum, the *Early Childhood Framework*, was first published in 1992 by the Project Construct National Center. The *Project Construct* approach is organized around 29 goals for students that are set within a context of four developmental domains:

- Cognitive;
- Representational;
- Sociomoral; and
- Physical.

The Project Construct National Center supports professional development through institutes, workshops, conferences, and on-site consultations as well as through extensive print and video materials.

Sample

The Missouri research team recruited full-day child-care centers through initial phone contacts followed by a letter to briefly explain the study. The program directors were asked to complete a “preschool information form” to clarify enrollment and demographics of the children and staff. If the data on the preschool information form appeared to meet the criteria for eligibility, the director was again contacted. Letters explaining the study and a cooperation agreement were sent to each director and teacher. The primary incentive was free training in *Project Construct* for the treatment group teachers in the initial year of the study and for the control teachers the following year. The treatment classrooms also received supplies and materials to support the implementation of *Project Construct*.

All of the preschools are full-day programs. The preschool program staff assisted with the recruitment of parents and children for the study. The average parental consent rate was 90 percent (90% for the treatment group, 89% for the control group). A total of 231 children and parents were recruited. Data were collected on a total sample of 228 children and 212 parents at the time of the fall baseline data collection.

In the follow-up year of the evaluation, the sample of schools went from 21 in pre-kindergarten to 124 schools in kindergarten. The sample of classrooms went from 23 preschool to 166 kindergarten classrooms. Data were collected on 188 children and 195 parents from the original sample.

Children and Families

The children were 4.7 years old at the time of baseline data collection and less than half (45%) were male. The majority of the sample of preschoolers was White (65%) or African American (25%). Table 12.1 provides additional information on the demographic characteristics of the children in the study sample. There were no statistically detectable differences between the treatment and control groups on the child characteristics.

Table 12.1. Child demographic characteristics for *Project Construct*

Characteristics	Full sample n = 231	Curriculum comparison	
		Control n = 108	Treatment n = 123
Age at baseline (years), mean	4.7	4.7	4.6
Gender (% male)	45.2	45.4	45.0
Race/ethnicity (%)			
White, non-Hispanic	64.8	66.0	63.7
African American, non-Hispanic	25.5	24.3	26.5
Hispanic	2.8	5.8	0.0
Asian or Pacific Islander	‡	‡	0.0
Native American	‡	‡	‡
Multiple/other	5.6	‡	8.0
Child disability status (parent reported, %)	15.1	13.0	17.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 12.2. The average age of the primary caregiver was 32 years. About half (52%) were married, and 27 percent were never married. More than half of the primary caregivers reported having some college (36%) or a bachelor's or higher (28%); 27 percent had a high school diploma or GED; and 8 percent had not finished high school. Most (74%) of the primary caregivers were employed full-time; 12 percent were employed part-time; and 12 percent were unemployed. At baseline, mothers in the treatment group were older relative to those assigned to the control group (33 years vs. 31 years, $p < .05$).

Teachers

There were 23 teachers who participated in the preschool year intervention study. All of the teachers were female, and most were White (70%) or African American (26%). On average, the preschool teachers had 10 years of teaching experience, with an average of 8 years of experience teaching preschool. The majority had no college education (61%) and 26 percent had a bachelor's degree. The majority (78%) reported having no teaching credential. Table 12.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 10.9 children. The child-staff ratio was on average 6.4 children to one teacher or program staff person.

Random Assignment

The Missouri research team identified and recruited a convenience sample of preschools from urban and rural locations in Missouri. Along with the Missouri researchers, Mathematica Policy Research, Inc. (MPR) determined the unit of random assignment for this research site. The MPR research staff randomly assigned preschool centers to treatment and control conditions because a preschool operated only one classroom or it was not feasible to vary the curriculum condition within a school. To increase the precision with which to estimate impacts, MPR grouped schools into blocks of two, and randomly assigned half the schools in each

Table 12.2. Primary caregiver demographic characteristics for *Project Construct*

Characteristics	Full sample n = 212	Curriculum comparison	
		Control n = 100	Treatment n = 112
Age at baseline (years), mean	32.2	31.0	33.3*
Marital status (%)			
Married	51.9	53.0	50.9
Separated/Divorced	19.8	16.0	23.2
Widowed	‡	‡	0.0
Never Married	27.4	29.0	25.9
Race/ethnicity (%)			
White, non-Hispanic	71.7	72.0	71.4
African American, non-Hispanic	25.0	23.0	26.8
Hispanic	2.4	4.0	‡
Asian or Pacific Islander	‡	‡	0.0
Native American	‡	0.0	‡
Multiple/other	0.0	0.0	0.0
Educational level (%)			
Did not finish high school	8.0	11.0	5.4
High school diploma or GED	27.4	25.0	29.5
Some college	36.3	42.0	31.3
College graduate	28.3	22.0	33.9
Employment (%)			
Full-time	74.1	73.0	75.0
Part-time	12.3	9.0	15.2
Unemployed	12.3	17.0	8.0
Other	‡	‡	‡

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

* $p < .05$

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

block to the treatment group and half to the control group. The MPR research staff formed blocks by matching schools on easily measured characteristics such as teachers' experience, school location, or score on a state report card system and, in doing so, increased the probability that those characteristics would be evenly distributed between the overall treatment and control groups. MPR staff used a random number function (RAND function in MS Excel) to generate random numbers. They sorted preschools by block and assigned a random number to each preschool. The preschools were then randomly assigned to treatment and control conditions. The staff assigned the highest ranking preschool within the block to the treatment condition, the next highest to the control condition, alternating assignment to treatment and control conditions until all preschools were randomly assigned to one of two conditions. Twenty-three preschool programs (26 preschool classrooms) were initially recruited and randomly assigned to treatment and control conditions. The final study sample of preschool programs included a total of 21 preschool centers (10 control and 11 treatment). The final sample of preschool centers included a sample of 23 preschool classrooms and teachers. There were a total of 11 control classrooms (one preschool center with two classrooms); and 12 treatment classrooms (one preschool center with two classrooms).

Table 12.3. Preschool teacher characteristics for *Project Construct*

Characteristics	Full sample n = 23	Curriculum comparison	
		Control n = 11	Treatment n = 12
Gender (% female)	100.0	100.0	100.0
Race/ethnicity (%)			
White, non-Hispanic	70.0	82.0	58.0
African American, non-Hispanic	26.0	‡	42.0
Hispanic	0.0	0.0	0.0
Asian or Pacific Islander	0.0	0.0	0.0
Native American	0.0	0.0	0.0
Multiple/other	‡	‡	0.0
Educational level (%)			
High school diploma or GED	61.0	55.0	67.0
Associate's degree	‡	0.0	‡
Bachelor's degree	26.0	36.0	‡
Graduate degree	0.0	0.0	0.0
Current teaching license/certificate (%)	‡	‡	‡
Child Development Associate (CDA) (%)	‡	‡	‡
State-awarded preschool certificate (%)	‡	‡	‡
No credential (%)	78.0	82.0	75.0
Years of teaching experience, overall (mean)	9.8	8.1	11.4
Years of preschool teaching experience (mean)	7.6	6.3	8.6

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Contamination

Because schools were assigned to either the treatment or the control group, there was little risk of contamination across the treatment and control conditions.

Control Condition

In the control schools, teacher-developed, generic curricula were implemented.

Data Collection

MPR collected the child, parent, teacher, and school data for the Missouri site for all three waves of data collection. The fall assessment data collection window for child assessments ranged from September 26, 2003 to November 11, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 42 days. The spring pre-kindergarten window was April 5, 2004 to June 20, 2004, and the kindergarten follow-up window was April 18, 2005 to June 8, 2005.

Attrition

A total of 26 classrooms/teachers (13 control and 13 treatment classrooms) were recruited at the beginning of the study. The final sample included 23 teachers and classrooms (11 control and 12 treatment classrooms) because two preschool programs (housing a total of three preschool classrooms) were dropped from the final study sample. One program (two classrooms) was closed and another program (one classroom) was folded into an existing preschool program because of low enrollment numbers. These changes resulted in a loss of

two preschool programs and three preschool classrooms (two control classrooms and one treatment classroom) in fall of the pre-kindergarten year.

For the child assessment, the fall 2003 response rate was 99 percent, the spring 2004 pre-kindergarten response rate was 90 percent, and the kindergarten follow-up response rate was 81 percent.

Implementation

The teachers who were assigned to the *Project Construct* treatment condition received training on three curriculum modules. The training consisted of three 12-hour modules; four 4-hour on-site consultations (the first three shortly after the completion of the modules) in the participants' classrooms; and two 3-hour follow-up workshops (attendance was not mandatory). The modules were designed for educators of children ages 3-5 years. The training on Module 1 (the young child and the learning environment) was held in August 2003; training on Module 2 (early literacy and the expressive arts) was held in October 2003; and training on Module 3 (young children's mathematical and scientific thinking) was held in November 2003. The three modules cover the entire *Project Construct* early childhood framework. The *Project Construct* training institute presented content that is aligned with the Missouri Pre-kindergarten Standards and covers levels 1-3 of the Missouri Core Teacher Competencies.

The onsite consultations occurred following the completion of each module training session. During the onsite consultations, the *Project Construct* consultant observed the teacher in her classroom during a regular classroom session. The consultant then provided the teacher with feedback based on topics from the previous module training and addressed how the training material could be incorporated into the teacher's practice. This feedback included changes in the classroom environment, curriculum planning, family involvement, and teacher-child interaction. Teachers discussed their strengths, areas of needs, goals, and questions or concerns with the consultants.

The *Project Construct* training institute offered follow-up workshops two times during the school year. The subject for each follow-up workshop was a topic identified by the module participants. The workshops provided opportunities for the participants to have interactive and in-depth experiences that provided learning strategies related to the identified topic as well as the opportunity to deepen their understanding of constructivism.

The Missouri research team collected site-specific curriculum fidelity data using the Project Construct Early Childhood Classroom Survey (PC-ECCOS). Observations were conducted in treatment and control classrooms in fall 2003, and spring 2004. The initial curriculum fidelity observation occurred September 19, 2003 to October 29, 2003. The second fidelity check occurred in April and May of 2004. The PC-ECCOS uses a three-point scale to measure evidence for curriculum implementation for each item (1 = *no evidence*; 2 = *some evidence*; 3 = *extensive evidence*). *No evidence* indicates that raters observed *no* evidence of constructivist activities/practices, which would suggest that *Project Construct* was not being implemented in a classroom. *Some evidence* indicates that the raters observed a *fair* amount of evidence for a constructivist approach and teachers are implementing *Project Construct* to some extent. *Extensive evidence* indicates that raters observed a classroom that is *exemplary* in its implementation of constructivism and *Project Construct*.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from "Not at All" (0) to "High" (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. The *Project Construct* curriculum was rated at the low Medium level (1.7) on the global

implementation fidelity measure. The control group curriculum was rated at the Medium level (2.3) on the global fidelity measure.

Impact Analysis Results

We begin with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data.

Project Construct—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-14a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-14a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 12.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences for the fall pre-kindergarten assessment.

There was no statistically detectable difference for the spring pre-kindergarten or kindergarten assessments on the WJ Applied Problems and the CMA-A Composite Score.

There was a statistically reliable negative effect on the Shape Composition scale ($ES = -.42, p < .05$) for spring pre-kindergarten assessments, such that students in the *Project Construct* classrooms had lower overall scores relative to students in the control classrooms. There was no statistically detectable difference in spring of the following year.

Based on the analyses for the three mathematics measures, we conclude that *Project Construct* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences for the fall pre-kindergarten assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three reading measures, we conclude that *Project Construct* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, (d)

race/ethnicity, (e) disability status as reported by parent, and (f) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Project Construct* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences for the fall pre-kindergarten assessment.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the two language measures, we conclude that *Project Construct* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]). The covariates were (a) child's age, (b) gender, (c) disability status as reported by the parent, (d) race/ethnicity, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Project Construct* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Project Construct—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-14b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-14b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 12.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically detectable difference between groups on the fall observation.

No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Project Construct* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

There were no statistically significant differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales for the spring pre-kindergarten observation.

Based on the analyses of the four Arnett scales, we conclude that *Project Construct* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales); (b) phonological awareness (TBRS Phonological Awareness scale); (c) language (TBRS Book Reading and Oral Language scales); and (d) early mathematics (TBRS Math Concepts scale) for the spring pre-kindergarten observation only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences between groups on any of the TBRS scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Project Construct* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Project Construct* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Project Construct* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Project Construct* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Project Construct*

The findings for *Project Construct* are summarized in table 12.4.

Table 12.4. Effect sizes for *Project Construct*

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.06	.08	—
CMA-A Mathematics Composite	-.11	-.06	—
Shape Composition ¹	-.42**	.12	—
Reading			
TERA	.00	-.03	—
WJ Letter Word Identification	-.05	.16	—
WJ Spelling	-.15	.00	—
Phonological awareness			
Pre-CTOPPP/CTOPP	.10	†	-.12
Language			
PPVT	.03	.10	—
TOLD	-.05	.01	—
Behavior			
SSRS Social Skills	.22	†	.12
SSRS Problem Behavior ²	-.08	†	.07
PLBS/LBS	.00	†	-.02
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	.54	—	
Teacher-child interaction			
Arnett Detachment ³	.12	—	
Arnett Harshness ³	-.13	—	
Arnett Permissiveness ³	-.02	—	
Arnett Positive Interactions	.46	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	.81	
TBRS Oral Language	†	.52	
TBRS Phonological Awareness	†	.01	
TBRS Print and Letter Knowledge	†	.34	
TBRS Written Expression	†	.43	
TBRS Math Concepts	†	.53	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

** $p < .01$ ¹ Building Blocks, Shape Composition task² Higher scores on this scale represent more negative child behaviors.³ Lower scores on this scale represent a more positive classroom environment.⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Page intentionally left blank.

Chapter 13. *Ready, Set, Leap!*: University of California, Berkeley (New Jersey site)

Curriculum

The University of California, Berkeley (California) researchers, in collaboration with RMC Research (RMC), implemented the *Ready, Set, Leap!* curriculum. *Ready, Set, Leap!* is a comprehensive, pre-kindergarten curriculum that combines research-based instructional approaches with multisensory technology. The curriculum is structured around 9 thematic units, each with 120 detailed lesson plans for large- and small-group instruction, and ongoing informal and formal assessment tools. The curriculum aligns with the goals and research requirements of the National Association for the Education of Young Children (NAEYC), the National Head Start Association, and the Early Reading First initiative. This balanced program stresses the importance of active and experiential learning, social and emotional development, teacher-child relationships, and the home-school connection.

All elements are incorporated into the curriculum to provide teachers with comprehensive pre-kindergarten instruction.

The curriculum topics include language and early literacy, mathematics, science, social studies, fine arts, health and safety, personal and social development, physical development, and technology applications.

The curriculum emphasizes the following elements:

- literacy and language development, focusing particularly on scaffolding;
- phonological awareness;
- alphabetic knowledge;
- print awareness;
- oral language development;
- reading aloud; and
- reading comprehension through story discussion.

The technology is designed within each thematic unit to provide center-based activities to integrate the senses of touch, sight, and sound by encouraging students to actively engage with literacy and language, and allowing students to have individualized feedback and support throughout the learning process. There is also a home component to encourage parent-child interactions that support children's learning activities in the preschool setting. The *Ready, Set, Leap!* program application includes family letters, take-home books, and specific strategies specifically for forging strong home-school connections.

Sample

The California research team recruited pre-kindergarten programs in New Jersey. Members of the research team attended a regional pre-kindergarten center meeting with the Director of Early Childhood programs in one large urban area and asked directors to contact them if they were interested in participating in the project. The research team then contacted individual centers that had NAEYC certification. All of the centers offered full-day academic pre-kindergarten programs, typically from about 9:00 a.m. to 3:30 p.m. The exact length of day varied because many of the centers have wrap-around services and children may arrive early in the morning and stay until early evening. From the pool of eligible centers, a total of 39 classrooms/ teachers

were recruited. Treatment and control group teachers and teaching assistants received an incentive for participating in the study.

The research team did not contact families directly but worked with the center directors and classroom teachers to recruit participants. The local site coordinators (i.e., members of the research team) worked with teachers and the director to obtain parental consent for parent and child participation in the study. Teachers asked the parents for their consent to participate in the study. The research team obtained informed consent for 470 parents and children. Approximately 89 percent of the eligible sample of parents and children agreed to participate in the study. The average parental consent rate was 89 percent (93% for the treatment group, 84% for the control group). A final sample of 286 parents and children (149 treatment, 137 control) were included in the study. Data were collected on a sample of 275 children and 261 parents at the time of the fall baseline data collection out of the final baseline sample of 286 parents and children.

In the follow-up year of the evaluation, participants from 21 preschools were followed into 94 schools in kindergarten. The sample of classrooms went from 39 preschool to 162 kindergarten classrooms. Data were collected on 248 children and 218 parents from the original sample.

Children and Families

The children were 4.5 years of age at the time of baseline data collection and more than half (54%) were male. The majority of the sample of preschoolers were African American (78%) or Hispanic (20%). Table 13.1 provides additional information on the demographic characteristics of the children in the study sample. There were no statistically detectable differences between the treatment and control groups on these child characteristics.

The demographic characteristics of the primary caregivers, who were most often the biological or adoptive mother, are presented in table 13.2. The average age of the primary caregiver was 31 years. Nearly two-thirds (63%) were never married, and about a quarter (26%) were married. Forty-three percent of the primary caregivers reported having a high school diploma or GED; 27 percent had some college; 12 percent had a bachelor's degree or higher; and 19 percent did not finish high school. More than half (53%) of the primary caregivers were employed full-time, 13 percent were employed part-time, and 32 percent were unemployed. There were no significant differences between the treatment and control groups on the primary caregiver characteristics.

Table 13.1. Child demographic characteristics for *Ready, Set, Leap!*

Characteristics	Full sample n = 286	Curriculum comparison	
		Control n = 137	Treatment n = 149
Age at baseline (years), mean	4.5	4.5	4.5
Gender (% male)	54.2	56.8	51.7
Race/ethnicity (%)			
White, non-Hispanic	0.0	0.0	0.0
African American, non-Hispanic	78.4	74.8	81.6
Hispanic	20.1	22.8	17.6
Asian or Pacific Islander	‡	‡	0.0
Native American	‡	‡	‡
Multiple/other	‡	‡	0.0
Child disability status (parent reported, %)	8.1	8.2	8.0

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Table 13.2. Primary caregiver demographic characteristics for *Ready, Set, Leap!*

Characteristics	Full sample n = 256	Curriculum comparison	
		Control n = 120	Treatment n = 136
Age at baseline (years), mean	30.8	30.9	30.8
Marital status (%)			
Married	26.2	25.8	26.5
Separated/Divorced	10.2	12.5	8.1
Widowed	‡	0.0	‡
Never Married	63.3	61.7	64.7
Race/ethnicity (%)			
White, non-Hispanic	‡	‡	0.0
African American, non-Hispanic	77.0	76.3	77.6
Hispanic	21.4	21.2	21.6
Asian or Pacific Islander	‡	‡	0.0
Native American	‡	‡	‡
Multiple/other	0.0	0.0	0.0
Educational level (%)			
Did not finish high school	18.6	21.2	16.3
High school diploma or GED	42.7	39.8	45.2
Some college	26.5	27.1	25.9
College graduate	12.3	11.9	12.6
Employment (%)			
Full-time	52.7	55.8	50.0
Part-time	12.9	13.3	12.5
Unemployed	32.4	27.5	36.8
Other	‡	3.3	‡

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Parent Interview (Fall 2003, Spring 2004, and Spring 2005).

Teachers

There were 39 teachers who participated in the preschool year intervention study. Most (95%) of the preschool teachers were female. The racial/ethnic composition of the sample included African American (61%), Hispanic (18%), and White (10%) teachers. The preschool teachers had an average of 8 years of teaching experience and 5 years of experience teaching preschool. The majority of teachers had a bachelor's degree (69%); 10 percent had an associate's degree; and 15 percent had no college degree. The majority of teachers reported having a current teaching license/certificate (51%), a state-awarded preschool certificate (49%), or a Child Development Associate (CDA) credential (26%). Table 13.3 provides additional information on the characteristics of the preschool sample of teachers. There were no statistically detectable differences between the treatment and control groups on the teacher characteristics.

Programs/Classrooms

The average preschool class size was 12.3 children. The child-staff ratio was an average of six children to one teacher or program staff person.

Table 13.3. Preschool teacher characteristics for *Ready, Set, Leap!*

Characteristics	Full sample n = 39	Curriculum comparison	
		Control n = 18	Treatment n = 21
Gender (% female)	95.0	94.0	95.0
Race/ethnicity (%)			
White, non-Hispanic	11.0	12.0	‡
African American, non-Hispanic	61.0	53.0	67.0
Hispanic	18.0	‡	19.0
Asian or Pacific Islander	‡	0.0	‡
Native American	0.0	0.0	0.0
Multiple/other	‡	‡	0.0
Educational level (%)			
High school diploma or GED	15.0	11.0	19.0
Associate's degree	10.0	17.0	‡
Bachelor's degree	69.0	72.0	67.0
Graduate degree	‡	0.0	‡
Current teaching license/certificate (%)	51.0	39.0	62.0
Child Development Associate (CDA) (%)	26.0	22.0	29.0
State-awarded preschool certificate (%)	49.0	39.0	57.0
No credential (%)	18.0	28.0	‡
Years of teaching experience, overall (mean)	8.0	6.8	9.0
Years teaching preschool (mean)	5.4	4.4	6.3

‡ Reporting standards not met. Values suppressed to protect participant confidentiality.

SOURCE: PCER Preschool Teacher Survey (Fall 2003 and Spring 2004).

Random Assignment

The California research team identified and recruited a convenience sample of 21 preschools from an urban area in New Jersey. Along with the California research team, Mathematica Policy Research, Inc. (MPR) determined the unit of random assignment at the research study site in New Jersey. The MPR research staff randomly assigned individual classrooms to conditions after it was determined that the experimental curriculum could be introduced in one classroom without affecting neighboring classrooms in the same school, and that preschool staff were willing to use different curricula within the same setting. To increase the precision with which to estimate impacts, MPR grouped classrooms into blocks of two or more and randomly assigned half the classrooms in each block to the treatment group and half to the control group. MPR research staff formed blocks by matching schools on easily measured characteristics such as teachers' experience, school location, or score on a state report card system and, in doing so, increased the probability that those characteristics would be evenly distributed between the overall treatment and control groups. MPR staff used a random number function (RAND function in MS Excel) to generate random numbers. They sorted the classrooms by block and assigned a random number to each classroom. The classrooms were then randomly assigned to treatment and control conditions. The staff assigned the highest ranking classroom within the block to the treatment condition, the next highest to the control condition, alternating assignment to treatment and control conditions until all classrooms were randomly assigned to one of two conditions. A total of 39 classrooms (21 treatment and 18 control) were randomly assigned to conditions. The 39 classrooms were drawn from 21 schools (10 schools and 18 control classrooms; 11 schools and 21 treatment classrooms).

Contamination

Because both *Ready Set Leap!* and control classrooms could reside within the same school, the researchers monitored the classrooms to ensure that treatment group teachers were not sharing materials and instructional practices with the control group teachers.

Control Condition

In the control condition, the teachers used the *High/Scope* approach to early childhood education.

Data Collection

MPR collected the child, parent, teacher, and school data for the New Jersey site for all three waves of data collection. The fall assessment data collection window for child assessments ranged from October 20, 2003 to November 19, 2003. The average delay from the beginning of the treatment (i.e., start of the school year) to the beginning of the fall assessment window was 35 days. The spring pre-kindergarten window was May 10, 2004 to June 15, 2004, and the kindergarten follow-up window was April 25, 2005 to June 8, 2005.

Attrition

Thirty-nine classrooms were randomly assigned to treatment and control conditions. All 39 classrooms remained in the study throughout the pre-kindergarten year.

For the child assessment, the fall 2003 response rate was 96 percent; the spring 2004 pre-kindergarten response rate was 92 percent; and the kindergarten follow-up response rate was 87 percent.

Implementation

The *Ready, Set, Leap!* curriculum was implemented in the 21 treatment classrooms in September 2003. Treatment group teachers received 4 full days of professional development training. The training sessions were scheduled to occur throughout the preschool year (September 2003, November 2003, January 2004, and March 2004). Curriculum fidelity was measured by triangulating three sources of data: (1) coaching visits that occurred three times during the school year; (2) site coordinator ratings based upon their three visits to each treatment and control classroom; and (3) modified CLASSIC observation coding based upon the site coordinator observations that included a 90-second time sampling procedure, with 32 events recorded for each observation, for a total of 117 observations. Inter-rater reliability was established for approximately 10 percent of the observations.

Implementation Fidelity Ratings

Each research team used a global fidelity measure to rate the overall fidelity with which the curricula were implemented in the preschool year of the project. A four-point scale ranging from “Not at All” (0) to “High” (3) was used to rate each treatment classroom. Researchers were asked to use their site-specific implementation and fidelity data to rate each treatment classroom on the global fidelity measure as High, Medium, Low, or Not at All. Researchers were also asked to provide a global rating for the control group curriculum. The *Ready, Set, Leap!* curriculum (1.9) and the control curriculum (2.0) were both rated at the Medium level on the global implementation fidelity measure.

Impact Analysis Results

We begin with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) and then present the analyses of the classroom observation data.

Ready, Set, Leap!—Child Outcomes

The unadjusted mean scores for child-level measures are reported in table C-15a in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-15a in appendix D. For all analyses of child-level measures, the following covariates were included: (a) child's age, (b) gender, (c) race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. The student-level effect sizes (ESs) are presented in table 13.4.

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). There were no statistically detectable differences between the treatment and control groups on the WJ Applied Problems for the fall pre-kindergarten, spring pre-kindergarten, and spring kindergarten assessments.

For the CMA-A Composite Score, there was no statistically detectable difference for the fall pre-kindergarten assessment. There was a statistically significant negative effect (ESs = $-.24$, $p < .05$) for the spring pre-kindergarten assessment, indicating that children in the *Ready, Set, Leap!* classrooms were outperformed by students in the control classrooms. There was no statistically detectable difference between the treatment and control groups for the spring kindergarten assessment.

For the Shape Composition scale, there was a statistically reliable difference favoring the *Ready, Set, Leap!* group on the fall assessment (ESs = $.25$, $p < .05$; follow-up analyses for this finding are included in appendix A). There was no statistically detectable difference between the treatment and control groups for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three mathematics measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. There were no statistically detectable differences for the fall assessment on these measures.

There were no statistically detectable differences on any of these measures for the spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the three reading measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

The phonological awareness measures were the Preschool Comprehensive Phonological and Print Processing (Pre-CTOPPP), Elision subtest, and the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest. We conducted a repeated measures analysis on the Pre-CTOPPP fall and spring pre-kindergarten data. There was no statistically detectable difference on the Pre-CTOPPP for the fall assessment.

There was no statistically detectable difference on the Pre-CTOPPP for the spring pre-kindergarten assessment.

We analyzed the kindergarten CTOPP data using analysis of covariance (ANCOVA). For the ANCOVA analysis, the covariates were the (a) Pre-CTOPPP fall assessment score, (b) child's gender, (c) age, race/ethnicity, (d) disability status as reported by parent, and (e) mother's education. There was no statistically detectable difference between groups on the CTOPP for the spring kindergarten assessment.

Based on the analyses of the Pre-CTOPPP and CTOPP, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] and Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. There were no statistically detectable differences on these measures for the fall pre-kindergarten assessment.

In the spring of the pre-kindergarten and kindergarten years, there were no statistically detectable differences between groups on either measure.

Based on the analyses of the two language measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

We conducted a repeated measures analysis for all three pre-kindergarten social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors [PLBS]). The covariates were (a) child's age, (b) gender, and (c) race/ethnicity, (d) disability status as reported by the parent, and (e) mother's education. There were no statistically detectable differences on these measures for the fall assessment.

For the spring pre-kindergarten assessment, there were no statistically detectable differences on any of these measures.

We analyzed the data from the kindergarten versions of the three behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) using analysis of covariance (ANCOVA). For the ANCOVA analyses, the covariates included (a) the fall pre-kindergarten score of the pre-kindergarten version of the relevant test, along with (b) child's age, (c) gender, (d) race/ethnicity, (e) disability status as reported by the parent, and (f) mother's education.

There were no statistically detectable differences between groups on any of these measures for the spring kindergarten assessment.

Based on the analyses of the three behavioral measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on children's social and learning behaviors relative to the control condition.

Ready, Set, Leap!—Classroom Outcomes

The unadjusted mean scores for classroom measures are reported in table C-15b in appendix C. Covariate adjusted mean differences and standard errors are reported in table D-15b in appendix D. For all analyses of classroom measures, the following variables were included in the model as covariates: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site. The classroom-level effect sizes (ES_c) are presented in table 13.4.

Overall classroom environment

We conducted a repeated measures analysis on the Early Childhood Environment Rating Scale-Revised (ECERS-R). There was no statistically significant difference between groups on the fall observation. No statistically detectable difference between groups was obtained for the spring pre-kindergarten observation.

Based on the analysis of the ECERS-R, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year, and conducted repeated measures analyses. There were no statistically detectable differences on these measures for the fall observation.

There were no statistically detectable differences between groups on the Arnett Detachment, Harshness, Permissiveness, or Positive Interaction scales for the spring pre-kindergarten observation.

Based on the analyses of the four Arnett scales, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

We obtained observations on classroom instruction in (a) early literacy (Teacher Behavior Rating Scale [TBRS] Print and Letter Knowledge and Written Expression scales); (b) phonological awareness (TBRS Phonological Awareness scale); (c) language (TBRS Book Reading and Oral Language scales); and (d) early mathematics (TBRS Math Concepts scale) for the spring pre-kindergarten assessment only. To analyze these data, ANCOVAs were conducted; the covariates were: (a) teacher has a BA degree, (b) previous teaching experience, (c) child/adult ratio in classroom, (d) average class size, (e) city size, and (f) geographic site.

There were no statistically detectable differences between groups on any of the TBRS scales.

Based on the analyses of the TBRS Print and Letter Knowledge and Written Expression scales, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on early literacy instruction relative to the control condition.

Based on the analysis of the TBRS Phonological Awareness scale, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on instruction in phonological awareness relative to the control condition.

Based on the analysis of the TBRS Book Reading and Oral Language scales, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on language instruction relative to the control condition.

Based on the analysis of the TBRS Math Concepts scale, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on early mathematics instruction relative to the control condition.

Summary of Findings for *Ready, Set, Leap!*

The findings for *Ready, Set, Leap!* are summarized in table 13.4.

Table 13.4. Effect sizes for *Ready, Set, Leap!*

Measure	Student-level effect sizes (ES _s)		
	RM analysis Spring Pre-K	RM analysis Spring K	ANCOVA Spring K
Mathematics			
WJ Applied Problems	.04	.00	—
CMA-A Mathematics Composite	-.24*	-.10	—
Shape Composition ¹	.08	.03	—
Reading			
TERA	.08	.01	—
WJ Letter Word Identification	.01	-.12	—
WJ Spelling	.20	.04	—
Phonological awareness			
Pre-CTOPPP/CTOPP	-.09	†	-.02
Language			
PPVT	.15	-.02	—
TOLD	-.11	-.03	—
Behavior			
SSRS Social Skills	-.05	†	-.03
SSRS Problem Behavior ²	-.03	†	.07
PLBS/LBS	.07	†	-.01
Measure	Classroom-level effect sizes (ES _c)		
	RM analysis Spring Pre-K	ANCOVA Spring Pre-K	
Global classroom quality			
ECERS-R	.16	—	
Teacher-child interaction			
Arnett Detachment ³	.19	—	
Arnett Harshness ³	.30	—	
Arnett Permissiveness ³	-.24	—	
Arnett Positive Interactions	.04	—	
Teacher instructional practices ⁴			
TBRS Book Reading	†	-.18	
TBRS Oral Language	†	-.24	
TBRS Phonological Awareness	†	.22	
TBRS Print and Letter Knowledge	†	-.02	
TBRS Written Expression	†	.10	
TBRS Math Concepts	†	-.10	

— Not available.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹ Building Blocks, Shape Composition task

² Higher scores on this scale represent more negative child behaviors.

³ Lower scores on this scale represent a more positive classroom environment.

⁴ ANCOVA models for the TBRS measures did not include baseline pretest scores because TBRS data were only collected in spring of the pre-kindergarten year.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Significance indications (p -values) in the table refer to the tests of contrasts between intervention and control groups that underlie the effect sizes reported here. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Page intentionally left blank.

References

- Abbott-Shim, M., and Sibley, A. (2001). *Assessment Profile for Early Childhood Programs: Research Edition II*. Atlanta, GA: Quality Counts, Inc.
- Abrams & Company (2000). *Let's Read with the Letter People*. Waterbury, CT: Abrams & Company.
- Administration for Children and Families (ACF) (2002a). A Descriptive Study of Head Start Families: *FACES Technical Report I*. Washington, DC: U.S. Department of Health and Human Services.
- Administration for Children and Families (ACF) (2002b). Head Start FACES 2000: A Whole-Child Perspective on Program Performance. Washington, DC: U.S. Department of Health and Human Services.
- Arnett, J. (1989). Caregivers in Day-Care Centers: Does Training Matter? *Journal of Applied Developmental Psychology, 10*: 541-552.
- Bradley, R.H., and Caldwell, B. (1984). The HOME Inventory and Family Demographics. *Developmental Psychology, 20*(2): 315-320.
- Bunce, B.H. (1995). *Building a Language-Focused Curriculum for the Preschool Classroom: Volume II*. Baltimore: Paul H. Brookes.
- Campbell, F.A., Ramey, C.T., Pungello, E.P., Sparling, J., and Miller-Johnson, S. (2002). Early Childhood Education: Young Adult Outcomes from the Abecedarian Project. *Applied Developmental Science, 6*: 42-57.
- Chaney, C. (1992). Language Development, Metalinguistic Skills, and Print Awareness in 3-year-old Children. *Applied Psycholinguistics, 13*: 485-514.
- Clements, D.H., and Sarama, J. (2003). DLM Early Childhood Express Math Resource Guide. Columbus, OH: SRA/McGraw-Hill.
- Clements, D., Sarama, J., and Liu, X. (in press). Development of a Measure of Early Mathematics Achievement Using the Rasch Model: The Research-Based Early Math Assessment. *Educational Psychology*.
- Cost, Quality, and Child Outcomes Study Team (1995). *Cost, Quality, and Child Outcomes in Child Care Centers, Technical Report*. University of Colorado at Denver, CO.
- Cunningham, A.E., and Stanovich, K.E. (1997). Early Reading Acquisition and its Relation to Reading Experience and Ability Ten Years Later. *Developmental Psychology, 33*: 934-945.
- Cunningham, A.E., Stanovich, K.E., and West, K.E. (1994). Literacy Environment and the Development of Children's Cognitive Skills. In E.M. Assink (Ed.), *Literacy Acquisition and Social Context* (pp. 70-90). Oxford, England: Harvester Wheatsheaf/Prentice Hall.

References

- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 39: 1-38.
- Dickinson, D.K., and Tabors, P.O. (2001). *Beginning Literacy with Language: Young Children Learning at Home and School*. Baltimore: Brookes Publishing.
- Dodge, D.T., Colker, L.J., and Heroam, C. (2002). *The Creative Curriculum for Preschool* (4th ed.). Washington, DC: Teaching Strategies, Inc.
- Downer, J., and Pianta, R. (2006). Academic and Cognitive Functioning in First Grade: Associations with Earlier Home and Child Care Predictors and with Concurrent Home and Classroom Experiences. *School Psychology Review*, 35: 11-30.
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L., Feinstein, L., Engel, A., Brooks-Gunn, J., Sexton, H., Duckworth, K., and Japel, C. (2006). School Readiness and Later Achievement. *Developmental Psychology*, 43(6): 1,428-1,446.
- Dunn, L., and Dunn, L. (1997). *Peabody Picture Vocabulary Test, Third Edition (PPVT-III): Form IIA*. Circle Pines, MN: American Guidance Service, Inc.
- Echols, L.D., West, R.F., Stanovich, K.E., and Zehr, K.S. (1996). Using Children's Literacy Activities to Predict Growth in Verbal Cognitive Skills: A Longitudinal Investigation. *Journal of Educational Psychology*, 88: 296-304.
- Ehri, L.C., and Wilce, L.S. (1980). The Influence of Orthography on Readers' Conceptualization of the Phonemic Structure of Words. *Applied Psycholinguistics*, 1: 371-385.
- Entwisle, D.R., and Alexander, K.L. (1988). Factors Affecting Achievement Test Scores and Marks of Black and White First Graders. *The Elementary School Journal*, 88: 449-471.
- Entwisle, D.R., and Alexander, K.L. (1992). Summer Setback: Race, Poverty, School Composition, and Mathematics Achievement in the First Two Years of School. *American Sociological Review*, 57: 72-84.
- Feise, R.J. (2002). Do Multiple Outcome Measures Require *p*-value Adjustment. *BMC Medical Research Methodology*, 2: 8-11.
- Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. New York: John Wiley & Sons.
- Gomez, E.V., Schaalje, G.B., and Fellingham, G.W. (2005). Performance of the Kenward-Roger Method When the Covariance Structure is Selected Using AIC and BIC. *Communications in Statistics: Simulation and Computation*, 34: 377-392.
- Gresham, F.M., and Elliott, S.N. (1990). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service, Inc.
- Harms, T., Clifford, R.M., and Cryer, D. (1998). *Early Childhood Environment Rating Scale. Revised Edition*. New York: Teachers College Press.

References

- Hart, B., and Risley, T.R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul H. Brookes Publishing Co.
- Harville, D.A. (1974). Bayesian Inference for Variance Components Using Only Error Contrasts. *Biometrika*, 61: 383-385.
- Harville, D.A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, 72: 320-338.
- Helm, J.H., and Katz, L. (2001). *Young Investigators: The Project Approach in the Early Years*. New York: Teachers College Press.
- Henderson, C. (1984). *Applications of Linear Models in Animal Breeding*. Ontario, Canada: University of Guelph.
- Henderson, C.R. (1990). Statistical methods in animal improvement: historical overview. In D. Gianola and K. Hammond (Eds.), *Advances in statistical methods for genetic improvement of livestock*. Berlin: Springer-Verlag.
- High Reach Learning (1997a). *High Reach Learning Curriculum for Pre-K: Butterflies*. Charlotte, NC: High Reach Learning.
- High Reach Learning (1997b). *High Reach Learning Curriculum for Pre-K: Framework*. Charlotte, NC: High Reach Learning.
- Hohmann, M., and Weikart, D.P. (2002). *Educating Young Children: Active Learning Practices for Preschool and Child Care Programs*. Ypsilanti, MI: High/Scope Press.
- Juel, C. (1988). Learning to Read and Write: A Longitudinal Study of 54 Children from First Through Fourth Grades. *Journal of Educational Psychology*, 78: 243-255.
- Juel, C. (1991). Beginning Reading. In R. Barr, M.L. Kamil, P.B. Mosenthal, and P.D. Pearson (Eds.), *Handbook of Reading Research*, 2: 759-788. Mahwah, NJ: Lawrence Erlbaum Associates.
- Katz, L.G., and Chard, S.C. (1989). *Engaging Children's Minds: The Project Approach*. Greenwich, CT: Ablex.
- Kenward, M.G., and Roger, J.H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, 53: 983-997.
- Keselman, H.J., Algina, J., Kowalchuk, R.K., and Wolfinger, R.D. (1998). A Comparison of Two Approaches for Selecting Covariance Structures in the Analysis of Repeated Measurements, *Communications in Statistics: Simulation and Computation*, 27: 591-604.
- Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole Publishing Company.
- Klein, A., and Starkey, P. (2000). *Child Math Assessment*. Berkeley, CA: Author.
- Klein, A., and Starkey, P. (2002). *Child Math Assessment-Abbreviated*. Berkeley, CA: Author.

References

- Klein, A., Starkey, P., and Ramirez, A.B. (2002). *Pre-K Mathematics Curriculum*. United States of America: Scott Foresman.
- Kreft, I.I., and de Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: Sage Publications.
- Landry, S.H., Crawford, A., Gunnewig, S., and Swank, P. (2002). *Teacher Behavior Rating Scale (TBR)*. Center for Improving the Readiness of Children for Learning and Education, unpublished research instrument.
- LeapFrog School House. (2003) *Ready, Set, Leap!* Emeryville, CA: Author.
- Lentz, F.E. (1988). Effective Reading Interventions in the Regular Classroom. In J.L. Graden, J.E. Zins, and M.J. Curtis (Eds.), *Alternative Educational Delivery Systems: Enhancing Instructional Options for All Students*. Washington, DC: National Association of School Psychologists.
- Lieberman, I.Y., Shankweiler, D., Fischer, F.W., and Carter, B. (1974). Explicit Syllable and Phoneme Segmentation in the Young Child. *Journal of Experimental Child Psychology*, 18: 201-212.
- Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS System for Mixed Models*, Cary, NC: SAS Institute Inc.
- Lonigan, C.J., and Farver, J.M. (2002) *Literacy Express: A Preschool Emergent Literacy Curriculum*. Tallahassee, FL: Author.
- Lonigan, C.J., Wagner, R.K., Torgesen, J.K., and Rashotte, C.A. (2002). *The Preschool Comprehensive Test of Phonological and Print Processing*. Tallahassee, FL: Florida State University.
- Mariner, C.L., Zaslow, M.J., and Sugland, B.W. (1998). *Factor Structure and Predictive Validity of the HOME-Short Form for Three Racial/Ethnic Groups in the National Longitudinal Survey of Youth-Child Supplement*. Methods Working Paper. Washington, DC: Child Trends, Inc.
- Mason, J.M., and Stewart, J. (1989). *CAP Early Childhood Diagnostic Instrument, (prepublication edition)*. Iowa City, IA: American Testronics.
- McCulloch, C.E., and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons.
- McDermott, P.A., Green, L.F., Francis, J.M., and Stott, D.H. (1999). *Learning Behaviors Scale*. Philadelphia: Edumetric and Clinical Science.
- McDermott, P.A., Green, L.F., Francis, J.M., and Stott, D.H. (2000). *Preschool Learning Behaviors Scale*. Philadelphia: Edumetric and Clinical Science.
- McDermott, P.A., Leigh, N.M., and Perry, M.A. (2002). Development and Validation of the Preschool Learning Behaviors Scale. *Psychology in the Schools*, 39(4): 353-36.
- McGrew, K.S., and Woodcock, R.W. (2001). *Woodcock Johnson Achievement Test-3rd Edition (WJ III). Technical Manual*. Itasca, IL: Riverside Publishing.

References

- Miles, S., and Stipek, D. (2006). Contemporaneous and Longitudinal Associations Between Social Behavior and Literacy Achievement in a Sample of Low-income Elementary School Children. *Child Development*, 77: 103-117.
- Missouri Department of Elementary and Secondary Education. (1992). Project Construct: A Framework for Curriculum and Assessment. Jefferson City, MO: Author.
- National Research Council. (2001). *Eager to Learn: Educating Our Preschoolers*. In B.T. Bowman, M.S. Donovan, and M.S. Burns (Eds.), Committee on Early Childhood Pedagogy. *Commission on Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.
- Natriello, G., McDill, E.L., and Pallas, A.M. (1990). *Schooling Disadvantaged Children: Racing Against Catastrophe*. New York: Teachers College Press.
- Newcomer, P.L., and Hammill, D.D. (1997). *Examiner's Manual. Test of Language Development (TOLD) Primary: Third Edition*. Austin, TX: Pro-Ed.
- Notari-Syverson, A., O'Connor, R.E., and Vadasy, P.F. (1998). *Ladders to Literacy: A Preschool Activity Book*. Baltimore: Paul H. Brookes Publishing Company.
- Patterson, H.D., and Thompson, R. (1971). Recovery of Interblock Information When Block Sizes are Unequal. *Biometrika*, 58: 545-554.
- Peisner-Feinberg, E.S., and Burchinal, M.R. (1997). Relations Between Preschool Children's Child-care Experiences and Concurrent Development: The Cost, Quality, and Outcomes Study. *Merrill-Palmer Quarterly*, 43: 451-477.
- Pellin, B., and Edmonds, E. (2001). *Bright Beginnings, 1997-2001*. Charlotte, NC: Charlotte-Mecklenburg Schools.
- Perfetti, C.A., Beck, I., Bell, L., and Hughes, C. (1987). Phonemic Knowledge and Learning to Read are Reciprocal: A Longitudinal Study of First Grade Children. *Merrill-Palmer Quarterly*, 33: 283-319.
- Perie, M., Grigg, W.S., and Donahue, P.L. (2005). *The Nation's Report Card: Mathematics 2005*. U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990-2005 Mathematics Assessments.
- Phelps, P.C. (2002). *Beyond Centers & Circle Time Curriculum*. Lewisville, NC: Kaplan Early Learning Company.
- Princiotta, D., Flanagan, K.D., and Germino-Hausken, E. (2006). *Fifth Grade: Findings From The Fifth-Grade Follow-up of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)* (NCES 2006-038) U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Radloff, L.S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1: 385-401.
- Reid, D.K., Hresko, W.P., and Hammill, D.D. (2001). *Test of Early Reading Ability-Third Edition*. Circle Pines, MN: American Guidance Service, Inc.

References

- Rice, M.L., and Wilcox, K.A. (1995). *Building a Language-Focused Curriculum for the Preschool Classroom: Volume I*. Baltimore: Paul H. Brookes.
- Rothman, K. (1990). No Adjustments are Needed for Multiple Comparisons. *Epidemiology*, 1: 43-46.
- Ruopp, R., Travers, J., Glantz, F.M., and Coelen, C. (1979). Children at the Center: Summary of Findings and Their Implications. *Final Report for the National Day Care Study, 1*. Cambridge, MA: Abt Associates.
- Savitz, D.A., and Olshan, A.F. (1995). Multiple Comparisons and Related Issues in the Interpretation of Epidemiologic Data. *American Journal of Epidemiology*, 142: 904-8.
- Scarborough, H.S. (1989). Prediction of Reading Disability from Familial and Individual Differences. *Journal of Educational Psychology*, 81(1): 101-108.
- Schaalje, G.B., McBride, J.B., and Fellingham, G.W. (2002). Adequacy of Approximations to Distributions of Test Statistics in Complex Mixed Linear Models. *Journal of Agricultural, Biological, and Environmental Statistics*. 7: 512-524.
- Schochet, P. (2005). *Statistical Power for Random Assignment Evaluations of Education Programs*. Princeton, NJ: Mathematica Policy Research, Inc.
- Schweinhart, L. (2004). *The High/Scope Perry Preschool Study Through Age 40: Summary, Conclusions, and Frequently Asked Questions*. Ypsilanti, MI: High/Scope Press.
- Searle, S.R., Casella, G.S., and McCulloch, C.E. (1992). *Variance Components*. New York: John Wiley & Sons.
- Shankweiler, D., Crain, S., Katz, L., and Fowler, A.E. (1995). Cognitive Profiles of Reading-disabled Children: Comparison of Language Skills in Phonology, Morphology, and Syntax. *Psychological Science*, 6(3): 149-156.
- Snow, C.E. (Ed.). (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Spybrook, J., Raudenbush, R.W., Liu, X., and Congdon, R. (2006). *Optimal Design for Longitudinal and Multilevel Research: Documentation for the "Optimal Design" Software*. Ann Arbor, MI: University of Michigan.
- SRA/McGraw-Hill (2003a). *DLM Early Childhood Express*. Desoto, TX: Author.
- SRA/McGraw-Hill (2003b). *Open Court Reading Pre-K*. Desoto, TX: Author.
- Stanovich, K.E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly*, 21: 360-407.
- Stone, J. (2002). *Animated Literacy*. La, Mesa, CA: Author. <http://www.animated-literacy.com>.
- Success for All Foundation. (2003a). *Curiosity Corner*. Baltimore: Author.
- Success for All Foundation. (2003b). *Kinder Corner*. Baltimore: Author.

References

- Swallow, W.H., and Monahan, J.F. (1985). Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. *Technometrics*, 28: 47-57.
- U.S. Department of Education, National Center for Education Statistics. (2006). Enrollment in Early Childhood Education Programs. *The Condition of Education 2006* (NCES 2006-071). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Health and Human Services. (2002). *A Descriptive Study of Head Start Families: FACES Technical Report I*. Washington, DC: Author.
- Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Wagner, R.K., Torgeson, J.K., and Rashotte, C.A. (1994). Development of Reading-related Phonological Processing Abilities: New Evidence of Bidirectional Causality from a Latent Variable Longitudinal Study. *Developmental Psychology*, 30: 73-87.
- Wagner, R.K., Torgeson, J.K., and Rashotte, C.A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.
- West, J., Denton, K., and Germino-Hausken, E. (2000). *America's Kindergarteners* (ECLS-K) (NCES 2000-070). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- West, J., Denton, K., and Reaney, L. (2001). *Children's Reading and Mathematics Achievement in Kindergarten and First Grade* (NCES 2002-125). U.S. Department of Education. National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Williams, K.T., and Wang, J.J. (1997). *Technical References to the Peabody Picture Vocabulary Test (PPVT-III)*. Circle Pines, MN: American Guidance Service, Inc.
- Wolfinger, R., Tobias, R., and Sall, J. (1994). Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models. *SIAM Journal on Scientific Computing* 15(6): 1294-1310.
- Wood, J. (2002). *Early Literacy and Learning Model*. Jacksonville, FL: Florida Institute of Education and the University of North Florida.
- Wright Group, McGraw Hill. (2001). *Doors to Discovery: A New Prekindergarten Program*. Bothell, WA: WrightGroup/McGraw-Hill.
- Zins, J., Bloodworth, M., Weissberg, R., and Walberg, H. (2004). The Scientific Base Linking Social and Emotional Learning to School Success. In J. Zins, R. Weissberg, M. Wang, and H. J. Walberg (Eds.), *Building Academic Success on Social and Emotional Learning: What Does the Research Say?* (pp. 1-22). New York: Teachers Press, Columbia University.

Appendix A: Secondary Analysis Results

Page intentionally left blank.

The Secondary Analysis presented in appendix A attempts to identify evidence of the possibility of early treatment effects and initial nonequivalence of the treatment and control groups. In addition, it presents additional analyses to examine whether the findings from the Main Analysis¹ would change if stricter criteria were applied to address nonequivalence.

Early Treatment Effects and Initial Nonequivalence

Early treatment effects could occur during the lag that occurred between the implementation of the curricula at the start of preschool and the baseline pre-testing. For 7 of the 12 research teams, the baseline pretesting began more than 2 weeks after the beginning of the curriculum implementation and for three teams the lag was 5 or more weeks. The models used in the Main Analysis cannot identify positive impacts the curricula might have had on student and classroom measures during this lag period. Therefore, they might underestimate the actual effects of the curricula.

Nonequivalence of the treatment and control group at the baseline pretest could be linked to an early treatment effect or to an unfortunate randomization. If the treatment and control groups were equivalent at the start of the year and the curricula did have a positive effect on the treatment group, then the treatment effect might appear in the baseline pretesting if there had been a long enough lag period. In this case, the treatment group might appear significantly differently than the control at the baseline due to the early treatment effect. Nonequivalence could also occur through an unfortunate randomization of the relatively small number of preschools or classrooms, compared to large scale-up studies, randomized for each curriculum evaluation. In this case, the treatment and control groups might not be equivalent from the start, and there would be relatively low power to detect the nonequivalence. If the treatment and control groups were initially different, then statistically significant differences in their mean post-test results might be due to their initial differences rather than to the impact of a curriculum.

To determine whether there was evidence of nonequivalence and early treatment, the results from the repeated measures models were used. The first step was to identify statistically significant differences between the treatment and control groups' baseline pretest means of each measure. If such a difference was found, the measure was extrapolated back to the beginning of the school year (the start of the treatment). The extrapolation procedure was based in the rate of growth in achievement found during the pre-kindergarten year using the time variable included in the repeated measures model (see appendix B for details). Using an assumption of linear growth over the preschool year (i.e., that the growth rate from the start of the year to the fall pretest was the same as the rate from the pretest to the post-test), the start of year values for the measure were estimated for the treatment and control groups based on their rates of achievement growth and the number of days in the lag period. These start-of-year measures were then statistically tested for equivalency.

If there was a significant difference at the baseline pretest but not at the start of the year, there is some evidence of an early treatment effect (i.e., the groups started out similarly at the beginning of the year but the treatment group made greater gains by the pretest). If there were significant differences at both the baseline pretest and the start-of-year, there is some evidence that the groups were nonequivalent to begin with. Table A-1 identifies the measures for each curriculum that show this type of evidence. The second column identifies any measures that were statistically significantly different at the baseline pretest. The third column notes whether those measures were statistically significantly different at the start of school. The fourth column identifies measures for which there is evidence of an early treatment effect (a significant difference at the pretest and no difference at the start of treatment), and the fifth identifies measures for which there is evidence of nonequivalence at baseline (a significant difference at both the start of treatment and the pretest).

¹ The term "Main Analysis" refers to the analyses presented in chapters 1-13. The term "Secondary Analysis" refers to the analyses presented in appendix A.

Out of the 255 measures examined (17 measures with multiple observations allowing them to be extrapolated backwards multiplied by the 15 curricula), 3 have some evidence of an early treatment effect and 11 have some evidence of nonequivalence.

Table A-1. Possible early treatment effects and non-equivalence at baseline

Curricula	Significant differences at baseline	Significant differences at start of school	Possible early treatment effect	Possible non-equivalence at baseline
<i>Bright Beginnings</i>	ECERS-R Arnett-D	Yes Yes		ECERS-R Arnett-D
<i>Creative Curriculum (Vanderbilt)</i>	ECERS-R Arnett-D	Yes No	Arnett-D	ECERS-R
<i>Creative Curriculum (UNC-Charlotte)</i>				
<i>Creative Curriculum with Ladders to Literacy</i>				
<i>Curiosity Corner</i>	SSRS Problem Behaviors Arnett-P	Yes Yes		SSRS Problem Behaviors Arnett-P
<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>	WJ Letter Word Identification TOLD	Yes No	TOLD	WJ Letter Word Identification
<i>Doors to Discovery</i>	TOLD Arnett-P	Yes Yes		TOLD Arnett-P
<i>Early Literacy and Learning Model</i>				
<i>Language-Focused Curriculum</i>				
<i>Let's Begin with the Letter People</i>	Arnett-P	Yes		Arnett-P
<i>Literacy Express</i>	WJ Letter Word Identification	Yes		WJ Letter Word Identification
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>	Shape Composition ¹	No	Shape Composition ¹	
<i>Project Approach</i>				
<i>Project Construct</i>				
<i>Ready, Set, Leap!</i>	Shape Composition ¹	Yes		Shape Composition ¹

¹ Building Blocks, Shape Composition task

NOTE: Arnett-D: Arnett Detachment scale

Arnett-P: Arnett Permissiveness scale

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Considering Initial Nonequivalence in the Analysis

To adjust for the possibility that some of the results of the Main Analysis were affected by initial nonequivalence, a Secondary Analysis was conducted. The Secondary Analysis analyzed the data in two ways. First, the same repeated measures models were used as in the Main Analysis but a stricter criterion was

applied to their results making use of the comparison of means at baseline and comparison of growth rates during pre-kindergarten. Second, ANCOVA models were estimated for all measures that had more than one observation (these had been estimated using repeated measures models in the Main Analysis) or a similar preschool baseline measure on a different scale. The earlier observation of the measure was used a covariate when estimating the program impact on the measure observed at a later time. This covariate helped control for possible differences in the measure between the treatment and control groups. The ANCOVA analyses act as a sensitivity analyses to determine whether similar results are obtained using an alternative modeling approach. Both sets of models included the same covariates that were used in the models for the Main Analysis.

Table A-2 identifies which models generated results for which measures. Column three identifies which type of repeated measures model was used and for which grades results were generated. Column four does the same for the ANCOVA models. The repeated measures model could be estimated only for those measures with at least two observations.

Table A-2. Secondary analysis: Outcomes, measures, models, and grades analyzed

Outcome	Measure	Repeated measures model	ANCOVA model with Pre-K baseline covariate
Reading	TERA	Spline: Pre-K and K	Pre-K and K
	WJ Letter Word Identification	Spline: Pre-K and K	Pre-K and K
	WJ Spelling	Spline: Pre-K and K	Pre-K and K
Phonological awareness ¹	Pre-CTOPPP	Simple: Pre-K	Pre-K
	CTOPP		K
Language	PPVT	Spline: Pre-K and K	Pre-K and K
	TOLD	Spline: Pre-K and K	Pre-K and K
Mathematics	WJ Applied Problems	Spline: Pre-K and K	Pre-K and K
	CMA-A	Spline: Pre-K and K	Pre-K and K
	Shape Composition ²	Spline: Pre-K and K	Pre-K and K
Pre-kindergarten behavior ¹	SSRS Social Skills	Simple: Pre-K	Pre-K
	SSRS Problem Behavior	Simple: Pre-K	Pre-K
	PLBS	Simple: Pre-K	Pre-K
Kindergarten behavior ¹	SSRS Social Skills		K
	SSRS Problem Behavior		K
	LBS		K
Classroom quality	ECERS-R	Simple: Pre-K	Pre-K
Teacher-child interaction	Arnett Detachment	Simple: Pre-K	Pre-K
	Arnett Harshness	Simple: Pre-K	Pre-K
	Arnett Permissiveness	Simple: Pre-K	Pre-K
	Arnett Positive Interaction	Simple: Pre-K	Pre-K

¹Pre-kindergarten and kindergarten measures are not on the same scale.

²Building Blocks, Shape Composition task

NOTE: The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten). Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Two criteria were then applied to the results. First, a new one (different for the repeated measures and ANCOVA models) was applied to determine which measures had enough evidence to be considered significant. Second, criteria similar to those used in the Main Analysis were applied to determine the findings on the five student-level outcomes and two classroom-level outcomes. These findings were compared to the findings from the Main Analysis to identify the possibility that nonequivalence may have affected the conclusions of this report.

Secondary Analysis Using Repeated Measures Models

In the Main Analysis, the repeated measures models (both spline and simple) provided a comparison of treatment and control means for the preschool post-test (for all the student-level measures and for six of the classroom-level measures) and a similar comparison for the kindergarten post-test (for eight of the student-level measures). The Secondary Analysis used these same results but in addition took advantage of two other results provided by the repeated measures models: the comparison of the baseline means and the comparison of the growth in achievement during pre-kindergarten. The comparison of the baseline means, if there was no statistically significant difference in the treatment and control group means, gave an initial indication that the groups were equivalent at the time of the pretest. The comparison of the growth in achievement during preschool, if there was a statistically significant greater average growth by the treatment group, provided additional assurance that any statistically significant difference in the post-test means of the treatment and control groups did not reflect initial nonequivalence.

Comparing achievement growth rates in treatment and control groups could only be used for the measures from pre-kindergarten. The repeated measures models tested the growth in achievement from fall pre-kindergarten to spring pre-kindergarten and from spring pre-kindergarten to spring kindergarten. The former tested the impact of the curricula on pre-kindergarten growth in achievement. The latter tested any difference in achievement growth from the end of pre-kindergarten to the end of kindergarten but it did not directly test the curricula's impact on this growth. As a result, the comparison of growth was only used in the Secondary Analysis for the pre-kindergarten results.

Using these additional results from the repeated measures model, the Secondary Analysis required three conditions to be met in order to conclude that a curriculum had a significant effect on a measure for pre-kindergarten: (1) no statistically detectable difference at the pre-kindergarten baseline assessment, (2) a statistically significant covariate-adjusted mean difference between groups at the spring pre-kindergarten post-test, and (3) a statistically significant difference in the rate of growth during pre-kindergarten between the treatment and control groups.

For kindergarten, the Secondary Analysis determined that a curriculum had a significant effect on a measure only if the following two conditions were met: (1) no statistically detectable difference in the pre-kindergarten baseline assessment, and (2) a statistically significant covariate-adjusted mean difference between groups at the spring kindergarten post-test. The lack of the growth comparison made the kindergarten analysis less conservative than the preschool analysis.

Secondary Analysis Using ANCOVA Models

In the Main Analysis, ANCOVA models were used with measures observed only one time. In some cases, similar measures on different scales were observed in pre-kindergarten and in kindergarten so that the pre-kindergarten measure could be included as a covariate in the kindergarten analysis of that measure (e.g., the Comprehensive Test of Phonological and Print Processing [CTOPP]). In the other cases no such covariate existed (e.g., the Teacher Behavior Rating Scale [TBR]) and the analysis could not control for the initial value of the measure. For the Secondary Analysis, ANCOVA models were used with all the measures for which a similar covariate could be included. This included any measures observed two or three times (which were analyzed with repeated measure models in the Main Analysis) and those measures observed only once but had a similar measure observed in pre-

kindergarten (the CTOPP and the three kindergarten behavior measures). Those measures with no similar covariate (the TBRS) were not included in the Secondary Analysis.

The ANCOVA models containing the pre-kindergarten baseline assessment covariate estimate expected means for a given measure at a single time point adjusted for the initial value of that measure. By including an initial value for a measure, the ANCOVA adjusted somewhat for any nonequivalence at the start of treatment although it could not adjust for any differential rates of growth in achievement that resulted from initial differences in the groups. For both pre-kindergarten and kindergarten, the Secondary Analysis concluded that a curriculum had a significant effect on a measure if a statistically significant difference was found in the covariate-adjusted post-test mean from the ANCOVA analyses.

Criteria to Determine Findings

Because of the number of statistical tests that were conducted, some results could be considered significant merely by chance. For example, eight statistical tests were conducted for each of the three reading and math results (start of treatment means, pre-kindergarten fall means, pre-kindergarten spring means, kindergarten spring means, rate of growth fall to spring pre-kindergarten, rate of growth spring pre-kindergarten to spring kindergarten, ANCOVA testing pre-kindergarten spring means, and ANCOVA testing kindergarten spring means) for a total of 24 statistical tests per subject. On average, with alpha at the .05-level, 1.2 tests could be statistically significant by chance. Similarly 16 statistical tests were conducted for the two language measures so 0.8 tests could be statistically significant by chance.

Moreover, within each of the outcomes (mathematics, reading, language, phonological awareness, and behavior) the measures were sufficiently intercorrelated (see table A-3) that an effect on one would not be expected to appear, except by chance, without indications of some effect on the others. Because of the number of tests that were conducted within an outcome and because the measures within an outcome were moderately correlated, criteria were used to decide if *the preponderance of evidence* supported a conclusion that the intervention curriculum resulted in a treatment effect on an outcome by spring of the pre-kindergarten year. These criteria were the same as those used in the Main Analysis.

In practice then, two sets of criteria were applied to the model results for the measures to determine the findings. The first determined whether a curriculum had an impact on a measure. The second determined whether a curriculum had an impact on the student or classroom-level outcomes made up of a group of measures. Table A-4 describes the two criteria. Columns 2 and 3 list the criteria used to determine whether a curriculum affected a measure using either the repeated measures model or the ANCOVA model. Column 4 lists the criteria used to determine whether a curriculum affected an outcome: it is the same criteria used in the Main Analysis.

Table A-3. Correlation matrix for student-level measures

Curricula	WJ	CMA-A	Shape Composition	WJ Letter Word TERA Identification	WJ Spelling	WJ Pre-CTOPPP	PPVT	TOLD	SSRS	SSRS	PLBS	
	Applied Problems	Mathematics Composite							Social Skills	Problem Behaviors		
Mathematics												
WJ Applied Problems	1.00	.67	.47	.67	.52	.48	.63	.70	.68	.21	-.11	.23
CMA-A Mathematics Composite	.67	1.00	.59	.61	.45	.48	.52	.52	.50	.09	-.09	.10
Shape Composition ¹	.47	.59	1.00	.38	.24	.41	.33	.32	.35	.06	-.05	.09
Reading												
TERA	.67	.61	.38	1.00	.70	.62	.46	.63	.55	.15	-.08	.16
WJ Letter Word Identification	.52	.45	.24	.70	1.00	.60	.54	.22	.34	.17	-.13	.16
WJ Spelling	.48	.48	.41	.62	.60	1.00	.29	.34	.31	.10	-.06	.13
Phonological awareness												
Pre-CTOPPP	.63	.42	.32	.54	.31	.29	1.00	.63	.60	.21	-.08	.11
Language												
PPVT	.70	.52	.32	.63	.22	.34	.63	1.00	.52	.19	-.06	.20
TOLD	.68	.50	.35	.55	.34	.31	.60	.70	1.00	.20	-.14	.16
Social skills												
SSRS Social Skills	.21	.09	.06	.15	.16	.10	.21	.19	.20	1.00	-.53	.66
SSRS Problem Behaviors	-.11	-.09	-.05	-.07	-.13	-.06	-.08	-.06	-.14	-.53	1.00	-.75
PLBS	.23	.10	.09	.16	.16	.13	.11	.20	.16	.66	-.75	1.00

¹ Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table A-4. Criteria used to determine curricula’s impact on a measure and on an outcome

Grade	Criterion 1: Determination that a curriculum affects a measure		Criterion 2: Determination that a curriculum affects an outcome
	Repeated measures model	ANCOVA model	
Pre-K	(a) No statistically significant difference in the preschool pretest means, and (b) a statistically significant covariate-adjusted mean difference at the preschool post-test, and (c) a statistically significant difference in the rate of growth during preschool	A statistically significant difference in the covariate-adjusted preschool post-test means	<p>For reading, math, and behavior, at least two of the three measures found to be positively affected (and none negatively)</p> <p>At least one of the two language measures found to be positively affected (and none negatively)</p> <p>The phonological awareness measure (Pre-CTOPPP) found to be positively affected</p> <p>The classroom quality measure (ECERS-R) found to be positively affected</p> <p>At least two of the four teacher-child interaction measures found to be positively affected (and none negatively)</p>

NOTE: ANCOVA: Analysis of covariance
 Pre-CTOPPP: Preschool Comprehensive Test of Phonological and Print Processing, Elision subtest
 ECERS-R: Early Childhood Environment Rating Scale-Revised
 SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Results of the Secondary Analysis

Appendix A contains a separate discussion of the Secondary Analysis for each curriculum with the following results (from the repeated measures models unless identified as from the ANCOVA models) provided in table form: (a) covariate adjusted-mean differences at the start of curriculum implementation; (b) covariate-adjusted mean differences at the time of the fall preschool baseline assessment; (c) covariate-adjusted mean differences at the time of the spring preschool post-test assessment (from both the repeated measures and ANCOVA models); (d) the fall to spring pre-kindergarten slope difference (rate of growth from fall to spring of pre-kindergarten); (e) the covariate-adjusted mean differences at the time of the kindergarten post-test assessment (from both the repeated measures and ANCOVA models); and (f) the spring pre-kindergarten to spring kindergarten slope difference (rate of growth from spring pre-kindergarten to spring of kindergarten).

These results are presented in effect size units. Cohen's *d* was used to provide a measure of the magnitude or size of the treatment effect. For effect sizes calculated at the classroom-level (ES_C), the effect size is the difference between the treatment and control classroom means divided by the pooled standard deviations for classrooms. Because the variation in measures taken at the classroom or group-level tends to be smaller than the variation in measures taken at the individual-level, effect sizes at the classroom-level are generally larger than effect sizes at the student-level. Cohen's *d* was also used to provide a measure of the slope effect sizes (ES_{Slope}). The slope effect size is the difference between the pre-kindergarten or kindergarten slopes for the treatment and control groups divided by the pooled standard deviations for the child or classroom measure of interest. The slope effect size is a measure of the difference in the rate of growth for the treatment and control groups. See appendix B for more details.

Before turning to the results for the individual curricula, tables A-5 and A-6 summarize the findings from the Secondary Analysis concerning the student and classroom-level outcomes and compare them with the findings from the Main Analysis. Table A-5 provides the findings on the student-level outcomes for the Main and Secondary Analyses and table A-6 provides the findings on the two classroom-level outcomes that could be analyzed under the Secondary Analysis (the four instructional outcomes could not be included). The results in the tables for the Secondary Analysis are footnoted with a "1" if from the repeated measures model and a "2" if from the ANCOVA model.

The tables show that the Secondary Analysis did not identify any curricula affecting the outcomes that were not already identified in the Main Analysis. Also, none of the curricula found to affect outcomes in the Main Analysis affected any additional outcomes under the Secondary Analysis.

The Secondary Analysis reduced the number of impacts found to occur, as would be expected from the application of stricter criteria. Table 5 shows that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* affected preschoolers' reading, phonological awareness, and language under the Main Analysis. Under the Secondary Analysis, it affected preschoolers' reading. *Project Approach* was found to have a negative effect on behavior in the Main Analysis and no effect on behavior in the Secondary Analysis. Table 6 shows that *Creative Curriculum* (UNC-Charlotte) had a positive impact on teacher-child interaction in the Main Analysis and no such effect in the Secondary Analysis. *Lets Begin with the Letter People* was found to have an impact on classroom quality in the Main Analysis but not in the Secondary Analysis.

The other findings from the Main Analysis are similarly found in the Secondary Analysis. *Curiosity Corner* had an effect on kindergarten reading in both analyses. The *Early Literacy and Learning Model (ELLM)* had an effect on language in both analyses. *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had an effect on kindergarten reading, phonological awareness, and language in both. *Pre-K Mathematics with DLM Early Childhood Express Math software* affected preschool mathematics under both analyses. *Creative Curriculum* (UNC-Charlotte) and *Literacy Express* both had positive impacts on classroom quality in both analyses.

Table A-5. Findings on student-level outcomes: Main and secondary analyses

Curricula	Main analysis					Secondary analysis				
	Reading	Phonological awareness	Language	Math	Behavior	Reading	Phonological awareness	Language	Math	Behavior
<i>Bright Beginnings</i>										
<i>Creative Curriculum (Vanderbilt)</i>										
<i>Creative Curriculum (UNC-Charlotte)</i>										
<i>Creative Curriculum with Ladders to Literacy</i>										
<i>Curiosity Corner</i>	Pre-K: 0 K: +					Pre-K: 0 K: + ¹				
<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>	Pre-K: + K: +	Pre-K: + K: +	Pre-K: + K: +			Pre-K: + ² K: 0	Pre-K: 0 K: + ²	Pre-K: 0 K: + ¹		
<i>Doors to Discovery</i>										
<i>Early Literacy and Learning Model</i>			Pre-K: 0 K: +					Pre-K: 0 K: + ¹		
<i>Language-Focused Curriculum</i>										
<i>Let's Begin with the Letter People</i>										
<i>Literacy Express</i>										
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>				Pre-K: + K: 0					Pre-K: + ¹ K: 0	
<i>Project Approach</i>					Pre-K: 0 K: -					
<i>Project Construct</i>										
<i>Ready, Set, Leap!</i>										

¹ Finding from repeated measures analysis.

² Finding from ANCOVA analysis.

NOTE: Abbreviations of the findings are:

Pre-K: Pre-kindergarten

K: Kindergarten

+: Finding of a positive impact

-: Finding of a negative impact

Blank cell: Finding of no impact

0: Finding of no impact (when an impact is found for the other grade)

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table A-6. Findings on classroom-level outcomes: Main and secondary analyses

Curricula	Main analysis		Secondary analysis	
	Classroom quality	Teacher-child interaction	Classroom quality	Teacher-child interaction
<i>Bright Beginnings</i>				
<i>Creative Curriculum (Vanderbilt)</i>				
<i>Creative Curriculum (UNC-Charlotte)</i>	+	+	+ ¹	
<i>Creative Curriculum with Ladders to Literacy</i>				
<i>Curiosity Corner</i>				
<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>				
<i>Doors to Discovery</i>				
<i>Early Literacy and Learning Model</i>				
<i>Language-Focused Curriculum</i>				
<i>Let's Begin with the Letter People</i>	+			
<i>Literacy Express</i>	+		+ ¹	
<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>				
<i>Project Approach</i>				
<i>Project Construct</i>				
<i>Ready, Set, Leap!</i>				

¹ Finding from ANCOVA analysis

NOTE: Abbreviations of the findings are:

+: Finding of a positive impact

Blank cell: Finding of no impact

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Bright Beginnings: **Vanderbilt University (Tennessee site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ESs) and slope effect sizes (ES_{Slope}) are presented in table A-7.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 8 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten scores.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the spring kindergarten assessment or the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included

the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the TERA, there was no statistically detectable difference between groups on (a) the covariate-adjusted means at the fall pre-kindergarten assessment, (c) the rate of growth from fall to spring, (d) the rate of growth from fall of pre-kindergarten to spring pre-kindergarten, or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. There was a statistically reliable difference in covariate-adjusted means between groups at the spring pre-kindergarten assessment ($E_S = .39, p < .05$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the TERA.

On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten or spring kindergarten means.

On the WJ Letter Word Identification test, there was (a) no statistically detectable difference in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on reading relative to the control condition.

Phonological Awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Tests of Phonological and Print Processing (Pre-CTOPPP). For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP) data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and the Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear

spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the PPVT, there were no statistically detectable differences between groups in the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups in the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the two language measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and the Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and the Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences between groups in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-7.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R there was a statistically detectable difference in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = 1.39, p < .05$). The statistically reliable difference on the ECERS-R scale score at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups ($ES_C = 1.52, p < .05$). However, there was not a statistically detectable difference between groups on (b) the spring pre-kindergarten observation, or on (c) the rate of change from the fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference for the spring pre-kindergarten observation.

Based on the analyses for the ECERS-R, we conclude that *Bright Beginnings* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there was a statistically detectable difference in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = -1.16, p < .05$). The statistically reliable difference on the Arnett Detachment scale score at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups ($ES_C = -1.47, p < .05$). However, there was no statistically detectable difference between groups for (b) the spring pre-kindergarten observation, or (c) the rate of change

from the fall to spring pre-kindergarten. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.²

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on classroom environment relative to the control condition.

Classroom instruction

Because the classroom instruction measures (TBRS Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in the spring pre-kindergarten observation, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Bright Beginnings*

The impact of *Bright Beginnings* on the child- and classroom-level measures is summarized in table A-7.

² Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups in the fall assessment, we did not obtain a statistically significant difference between groups in the spring pre-kindergarten assessment.

Table A-7. Secondary analysis results for *Bright Beginnings*

Measure	RM analysis	RM analysis	RM analysis	Fall-Spring	ANCOVA ²	RM analysis	Spring Pre-K-	ANCOVA
	start of treatment ¹	Fall Pre-K	Spring Pre-K	slope	Spring Pre-K	kindergarten	Spring K slope	kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	-.08	-.04	.16	.1931	.18	.13	-.0173	.11
CMA-A Mathematics Composite	-.06	-.02	.14	.1595	.10	.07	-.0420	-.01
Shape Composition ³	.11	.09	-.03	-.1162	-.07	.15	.0958	.08
Reading								
TERA	.02	.09	.39*	.2908	.32	-.07	-.2478	-.19
WJ Letter Word Identification	.24	.26	.35	.0912	.11	.09	-.1426	-.08
WJ Spelling	-.02	.02	.18	.1581	.20	.06	-.0665	-.02
Phonological awareness								
Pre-CTOPPP/CTOPP	.07	.04	-.07	-.1130	-.08	†	†	.01
Language								
PPVT	-.10	-.05	.13	.1815	.12	.07	-.0345	.04
TOLD	-.15	-.10	.09	.1932	.18	.16	.0344	.14
Behavior								
SSRS Social Skills	-.35	-.33	-.27	.0643	-.12	†	†	-.03
SSRS Problem Behavior ⁴	.04	.07	.23	.1493	.19	†	†	.24
PLBS/LBS	.05	.04	.04	-.0059	.03	†	†	-.30
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	1.52*	1.39*	.80	-.5726	1.53	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	-1.47*	-1.16*	.19	1.3204	.15	†	†	†
Arnett Harshness ⁵	-.85	-.67	.12	.7694	-.04	†	†	†
Arnett Permissiveness ⁵	-.61	-.47	.16	.6148	.10	†	†	†
Arnett Positive Interactions	.96	.86	.41	-.4368	-.14	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Creative Curriculum: **Vanderbilt University (Tennessee site)**

Creative Curriculum was evaluated by two research teams—Vanderbilt University (Tennessee) and University of North Carolina at Charlotte (North Carolina). Here we present analyses from the Tennessee site, beginning with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-8.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 8 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition task). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups on the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups on the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that *Creative Curriculum* did not have statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's

education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were *fall assessment score*, child age, gender, disability status as reported by parent, race/ethnicity, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted for the pre-kindergarten Pre-CTOPPP data and the kindergarten CTOPP data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

On the ANCOVA for the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the phonological awareness measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture and Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included).

In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) spring kindergarten assessment, but there was a statistically detectable difference on the (d) rate of growth favoring the *Creative Curriculum* group from fall pre-kindergarten to spring pre-kindergarten (difference in slope = 4.25; $ES_{\text{slope}} = .2414$, $p < .01$). We did not obtain all three conditions necessary to indicate statistical evidence of a treatment effect on the PPVT. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted means for the spring pre-kindergarten assessment.

For kindergarten, on the PPVT, there were no statistically detectable differences between groups on the (e) spring kindergarten assessment or on the (f) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring kindergarten.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

Based on the analyses of the two language measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills, SSRS Problem Behaviors, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from

fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses for the behavioral measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{slope}) are presented in table A-8.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there was statistically detectable difference in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_c = 1.94, p < .01$). However, there was no statistically detectable difference between groups for the (b) spring pre-kindergarten observation, (c) the spring kindergarten observation, or (d) the rate of change from the fall to spring observation. The statistically reliable difference on the ECERS-R scale score at the fall observation suggests either the nonequivalence of the treatment and control groups or early implementation of the study curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups ($ES_c = 2.28, p < .001$). However, there was not a statistically detectable difference between groups for (b) the spring pre-kindergarten observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.³

Based on the analyses for the ECERS-R, we conclude that *Creative Curriculum* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences between the groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, the (b) spring pre-kindergarten

³ Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.⁴

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of change from the fall to spring assessment. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten assessment.

Based on the analyses for the teacher-child relationships measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because data derived from the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in the spring pre-kindergarten assessment, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Creative Curriculum* (Tennessee site)

The impact of *Creative Curriculum* on the child- and classroom-level measures is summarized in table A-8.

⁴ Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-8. Secondary analysis results for *Creative Curriculum*: Tennessee

Measure	RM analysis	RM analysis	RM analysis	Fall-Spring	ANCOVA ²	RM analysis	Spring Pre-K-	ANCOVA
	start of treatment ¹	Fall Pre-K	Spring Pre-K	slope	Spring Pre-K	kindergarten	Spring K slope	kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.09	.10	.17	.0671	.07	.17	-.0013	.09
CMA-A Mathematics Composite	.05	.06	.10	.0363	.03	.05	-.0265	.04
Shape Composition ³	.13	.09	-.12	-.2036	-.13	.00	.0629	-.04
Reading								
TERA	-.08	-.06	.02	.0791	.06	.10	.0430	.03
WJ Letter Word Identification	.32	.29	.16	-.1281	-.11	.38	.1156	.08
WJ Spelling	-.12	-.06	.19	.2402	.20	.25	.0345	.21
Phonological awareness								
Pre-CTOPPP/CTOPP	-.16	-.15	-.10	.0406	-.01	†	†	.06
Language								
PPVT	-.07	-.01	.23	.2414**	.21	.12	-.0624	.08
TOLD	-.07	-.05	.07	.1109	.09	.11	.0234	.14
Behavior								
SSRS Social Skills	-.27	-.22	-.03	.1847	.09	†	†	.35
SSRS Problem Behavior ⁴	.01	.02	.07	.0544	.05	†	†	-.05
PLBS/LBS	-.03	.00	.14	.1357	.13	†	†	.08
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	2.28**	1.94**	.45	-1.4470	1.57	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	-1.13	-.95*	-.16	.7686	-.16	†	†	†
Arnett Harshness ⁵	-.36	-.32	-.12	.1945	-.53	†	†	†
Arnett Permissiveness ⁵	-.27	-.13	.51	.6173	.60	†	†	†
Arnett Positive Interactions	.95	.74	-.15	-.8677	-.50	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Creative Curriculum: University of North Carolina at Charlotte (North Carolina and Georgia sites)

Creative Curriculum was evaluated by the University of North Carolina research team and by the Vanderbilt University research team. Here we present the results of the North Carolina research team evaluation. The student-level effect sizes (ES_s) and slope effect sizes (ES_{slope}) are presented in table A-9.

The North Carolina team implemented *Creative Curriculum* at sites in North Carolina and in Georgia. We present the analyses that combine the two implementation sites. We begin with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 16 days (including Saturdays, Sundays, and holidays) for North Carolina and 14 days for Georgia.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was a statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means ($ES_s = .32, p < .05$), and no statistically detectable significant differences between groups in the covariate-adjusted spring kindergarten means. In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on Shape Composition at spring pre-kindergarten relative to the control condition.

Based on the analyses of the three mathematics measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically significant differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means.

On the ANCOVA for the CTOPP, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture and Vocabulary Test [PPVT] and Test of Early Learning Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the language measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors scale measure, there were no statistically detectable differences in covariate-adjusted means at the (a) fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Creative Curriculum* did not have a statistically detectable effect on behavior relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-9.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant difference between groups on the (b) spring pre-kindergarten observation ($ES_C = 1.66, p < .05$). On the ANCOVA, a statistically significant difference was obtained on the spring pre-kindergarten observation ($ES_C = 1.36, p < .01$). *Creative Curriculum* classrooms received higher global classroom quality ratings relative to the control group classrooms.

Based on the analyses of the ECERS-R, we conclude that *Creative Curriculum* had a positive effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant difference between groups on (b) the spring pre-kindergarten observation ($ES_C = -1.68, p < .05$). *Creative Curriculum* teachers were rated as less detached in their interactions with students relative to teachers in the control classrooms. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there was no statistically significant difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. There was a statistically significant difference between groups on the (b) spring pre-kindergarten observation ($ES_c = 1.65, p < .01$). Teachers in *Creative Curriculum* classrooms were more positive in their interactions with students relative to teachers in the control classrooms as measured by the Arnett Positive Interactions scale. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship scales, we conclude that *Creative Curriculum* had a positive effect on teacher-child relationships relative to the control condition in pre-kindergarten but no effect in kindergarten.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Creative Curriculum* (North Carolina and Georgia sites)

The impact of *Creative Curriculum* on the child- and classroom-level measures is summarized in table A-9.

Table A-9. Secondary analysis results for *Creative Curriculum*: North Carolina and Georgia

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K- Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.17	.17	.20	.0273	.16	.09	-.0579	.08
CMA-A Mathematics Composite	.04	.01	-.10	-.1125	-.05	.14	.1273	.16
Shape Composition ³	-.03	.01	.19	.1703	.32*	-.01	-.1036	.03
Reading								
TERA	.17	.13	-.08	-.1968	-.20	-.04	.0194	-.16
WJ Letter Word Identification	-.22	-.19	-.08	-.1131	.05	.00	.0392	.16
WJ Spelling	.06	.01	-.18	-.1868	-.22	-.05	.0688	.02
Phonological awareness								
Pre-CTOPPP/CTOPP	.00	.00	.02	.0144	.05	†	†	.06
Language								
PPVT	.00	.02	.08	.0595	.11	.15	.0377	.12
TOLD	.21	.14	-.16	-.2884	-.17	-.17	-.0095	-.25
Behavior								
SSRS Social Skills	.22	.19	.05	-.1375	.00	†	†	-.12
SSRS Problem Behavior ⁴	-.10	-.11	-.16	-.0468	-.13	†	†	.08
PLBS/LBS	.21	.18	.07	-.1109	.02	†	†	-.20
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	.33	.58	1.66*	1.0578	1.36**	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	-.40	-.64	-1.68*	-1.0160	-1.25	†	†	†
Arnett Harshness ⁵	-.66	-.67	-.70	-.0260	-.18	†	†	†
Arnett Permissiveness ⁵	.67	.35	-1.01	-1.3300	-.76	†	†	†
Arnett Positive Interactions	.15	.43	1.65**	1.1926	1.40	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Creative Curriculum with Ladders to Literacy: **University of New Hampshire (New Hampshire site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ESs) and slope effect sizes (ES_{slope}) are presented in table A-10.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 10 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten or kindergarten means.

For the CMA-A Composite Score, there were no statistically significant differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) the rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on (a) the fall or (b) the spring pre-kindergarten assessments, but there was a statistically reliable difference on (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth =12.9; $ES_{\text{Slope}} = .5228, p < .05$) favoring the treatment group. We did not obtain all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Spelling. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means.

For kindergarten, on the WJ Spelling test, there were no statistically detectable differences between groups in (d) covariate-adjusted means for the spring kindergarten assessments or in the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity,

disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on (a) the fall assessment or (b) the spring pre-kindergarten assessment, but there was a statistically significant difference on the (c) rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = -6.2; $ES_{\text{slope}} = -.3262$, $p < .05$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the PPVT. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for the spring pre-kindergarten assessment.

For kindergarten, on the PPVT, there was no statistically detectable difference in (d) covariate-adjusted means at the spring kindergarten or the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in covariate-adjusted means for spring kindergarten.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the two language measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on language development relative to the control condition.⁵

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating scale [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills, SSRS Problem Behaviors, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for the spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no

⁵ The statistically significant difference between groups in rates of growth from pre-kindergarten spring to kindergarten spring does not "count" as a statistically significant test supporting a kindergarten effect because this slope does not address the impact of the intervention.

statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted, and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Ladders for Literacy* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-10.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or in (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained in the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Creative Curriculum with Ladders to Literacy* did not have a statistically detectable effect on overall classroom quality.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Creative Curriculum with Ladders to Literacy*

The impact of *Creative Curriculum with Ladders to Literacy* on the child- and classroom-level measures is summarized in table A-10.

Table A-10. Secondary analysis results for *Creative Curriculum with Ladders to Literacy*

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K- Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	-.15	-.15	-.14	.0072	.03	-.33	-.0997	-.28
CMA-A Mathematics Composite	.11	.12	.18	.0510	.11	-.19	-.1960	-.28
Shape Composition ³	-.08	-.07	.02	.0820	.10	-.10	-.0633	-.11
Reading								
TERA	.18	.09	-.30	-.3784	-.30	-.54	-.1271	-.60*
WJ Letter Word Identification	-.07	-.09	-.16	-.0734	.04	-.27	-.0585	-.17
WJ Spelling	-.36	-.24	.30	.5228*	.27	-.08	-.2009	-.23
Phonological awareness								
Pre-CTOPPP/CTOPP	.00	-.03	-.16	-.1262	-.12	†	†	-.10
Language								
PPVT	.03	-.04	-.38	-.3262*	-.22	-.30	.0438	-.29
TOLD	.04	-.01	-.22	-.2046	-.17	-.06	.0843	-.02
Behavior								
SSRS Social Skills	-.28	-.28	-.25	.0250	-.06	†	†	.17
SSRS Problem Behavior ⁴	-.03	-.02	-.01	.0178	-.02	†	†	.02
PLBS/LBS	-.20	-.18	-.08	.0991	-.03	†	†	-.11
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	.86	.57	-.71	-1.2460	-.07	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	-.42	-.24	.51	.7399	-.02	†	†	†
Arnett Harshness ⁵	.85	.64	-.26	-.8805	-.07	†	†	†
Arnett Permissiveness ⁵	.12	.29	1.02	.7151	.67	†	†	†
Arnett Positive Interactions	.67	.55	.03	-.5041	.99	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

²The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³Building Blocks, Shape Composition task

⁴Higher scores on this scale represent more negative child behaviors.

⁵Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Curiosity Corner: **Success for All Foundation (Kansas, Florida, and New Jersey sites)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The Success for All (SFA) team implemented its evaluation in three separate sites. Our discussion of the results focuses on the combined analyses of the three sites. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-11.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 14 days (including Saturdays, Sundays, and holidays) in Kansas, 35 days in New Jersey, and 49 days in Florida.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three mathematics measures, we conclude that *Curiosity Corner* did not have an effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's

education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the TERA, there were no statistically detectable differences between groups in the covariate adjusted means at the (a) fall pre-kindergarten assessment (b) spring pre-kindergarten assessment, or (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means.

For kindergarten, on the TERA, there was (d) a statistically significant difference on the spring kindergarten assessment ($ES_S = .43, p < .05$) and (e) a statistically significant difference in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = .05; $ES_{Slope} = .1771, p < .05$). On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically significant differences in covariate-adjusted means at the (a) fall, or (b) spring pre-kindergarten assessments, and (c) no difference in the rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means.

For kindergarten, on the WJ Letter Word Identification test, there was a statistically significant difference in the spring kindergarten assessment ($ES_S = .43, p < .05$) and a statistically significant difference in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = 4.74; $ES_{Slope} = .1806, p < .05$). On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically significant differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or kindergarten assessments, and no statistically significant differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically significant differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on pre-reading skills at the end of pre-kindergarten. However, *Curiosity Corner* had a positive effect on reading relative to the control condition at the end of the kindergarten year.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted of the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture and Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding scale) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA Grammatical Understanding scale, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the language measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating Scale [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there was a statistically significant difference between groups on the (a) fall pre-kindergarten assessment ($ESs = .53, p < .05$). There were no statistically detectable differences between groups on the (b) spring pre-kindergarten assessment, or the (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. The statistically reliable difference in Problem Behavior scores at the fall assessment suggests either the nonequivalence of treatment or control groups or an early treatment effect. To examine the possibility of an early treatment effect, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups on the Problem Behaviors measure ($ESs = .56, p < .05$). This finding suggests, but does not prove, nonequivalence at the start of treatment. On

the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment or (b) spring pre-kindergarten assessment, and no statistically detectable difference in (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Curiosity Corner* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_{slope}) and slope effect sizes (ES_{slope}) are presented in table A-11.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Curiosity Corner* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, there was no statistically significant difference between groups on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there was a statistically significant difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = -1.46, p < .05$). There was no statistically detectable difference on the (b) spring pre-kindergarten observations or the (c) rate of change from the fall to spring observation. The statistically reliable difference in Arnett Permissiveness scores at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the study curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the treatment group on the Permissiveness measure ($ES_C = -1.57, p < .05$). On the ANCOVA, no statistically significant difference was obtained on the spring pre-kindergarten observation.⁶

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship scales, we conclude that *Curiosity Corner* did not have a statistically detectable effect on the teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Curiosity Corner*

The impact of *Curiosity Corner* on the child- and classroom-level measures is summarized in table A-11.

⁶ Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-11. Secondary analysis results for *Curiosity Corner*

Measure	RM analysis	RM analysis	RM analysis	Fall-Spring	ANCOVA ²	RM analysis	Spring Pre-K-	ANCOVA
	start of treatment ¹	Fall Pre-K	Spring Pre-K	slope	Spring Pre-K	kindergarten	Spring K slope	kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.06	.06	.10	.0318	.10	.26	.0885	.06
CMA-A Mathematics Composite	.01	.01	.01	.0001	-.01	-.05	-.0330	.03
Shape Composition ³	-.11	-.06	.16	.2143	.07	.32	.0901	.41
Reading								
TERA	.33	.29	.10	-.1816	-.06	.43*	.1771*	.32
WJ Letter Word Identification	.26	.23	.09	-.1328	-.02	.43*	.1806*	.29
WJ Spelling	-.16	-.12	.04	.1515	.05	.20	.0906	.19
Phonological awareness								
Pre-CTOPPP/CTOPP	-.06	-.01	.18	.1866	-.01	†	†	.25
Language								
PPVT	-.04	-.04	-.01	.0273	-.04	.14	.0785	.17
TOLD	-.02	-.03	-.08	-.0409	-.05	.15	.1198	.15
Behavior								
SSRS Social Skills	-.26	-.23	-.06	.1598	-.10	†	†	.32
SSRS Problem Behavior ⁴	.56*	.53*	.43	-.1056	.07	†	†	-.08
PLBS/LBS	-.43	-.40	-.25	.1425	.02	†	†	.11
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	-.69	-.65	-.48	.1661	-.36	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	-.07	-.14	-.41	-.2668	1.40	†	†	†
Arnett Harshness ⁵	.34	.30	.14	-.1564	1.08	†	†	†
Arnett Permissiveness ⁵	-1.57*	-1.46*	-.98	.4708	-.60	†	†	†
Arnett Positive Interactions	.72	.59	.02	-.5506	-1.43	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Doors to Discovery: **University of Texas Health Science Center at Houston (Texas site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ESs) and slope effect sizes (ES_{slope}) are presented in table A-12.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 20 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). Each model included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (fall assessment score was not included). For each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, or between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

There was no statistically detectable difference between groups in (d) the spring kindergarten assessment, but there was a statistically reliable difference between groups in the (e) rates of growth from spring pre-kindergarten to spring kindergarten (difference in rates of growth = $-.04$; $ES_{\text{slope}} = -.1551$, $p < .05$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three mathematics measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included

the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) the rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups in the covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity,

disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences in covariate-adjusted means at the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatical Understanding subtest, there (a) was a statistically detectable difference between groups on the fall assessment favoring the treatment group ($ES_s = .38, p < .05$), but no statistically detectable differences (b) at the spring pre-kindergarten assessment, or (c) in the rate of growth from fall pre-kindergarten to spring pre-kindergarten. A statistical difference at the fall assessment could reflect either the failure of randomization to create equivalent groups or an early treatment effect. Extrapolating back to the beginning of the school year, we found a statistically reliable difference between groups favoring the *Doors to Discovery* group ($ES_s = .42, p < .05$). This finding suggests, but does not prove, nonequivalence at the start of treatment. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten. There were no statistically detectable differences (d) at the spring kindergarten assessment or (e) in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the two language measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the Preschool Learning Behaviors Scale, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the Learning Behaviors Scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the four behavioral measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{slope}) are presented in table A-12.

Overall classroom environment

We obtained observations on the ECERS-R in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically significant difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Doors to Discovery* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationships measures, ANCOVAs were conducted with the following covariates: *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there was a statistically significant difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_c = 1.06, p < .05$). However, there were no statistically detectable differences between groups on the (b) spring pre-kindergarten observation, or

(c) the rate of change from the fall to spring observation. The statistically reliable difference in Arnett Permissiveness scores at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the study curriculum. To examine the possibility of early implementation of the study curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the treatment group on the Permissiveness measure ($ES_c = 1.28, p < .05$). *Doors to Discovery* teachers were more permissive in their interactions with students relative to teachers in the control classrooms as measured by the Arnett Permissiveness scale. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.⁷

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of teacher-child relationship measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Doors to Discovery*

The impact of *Doors to Discovery* on the child- and classroom-level measures is summarized in table A-12.

⁷ Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-12. Secondary analysis results for *Doors to Discovery*

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.13	.11	.01	-.0932	.00	-.02	-.0185	-.05
CMA-A Mathematics Composite	.03	.05	.13	.0816	.16	-.16	-.1551*	-.15
Shape Composition ³	-.11	-.11	-.13	-.0155	-.13	-.12	.0021	-.07
Reading								
TERA	.18	.16	.06	-.1004	-.05	-.05	-.0586	-.15
WJ Letter Word Identification	.10	.10	.10	-.0042	.09	-.09	-.0993	-.14
WJ Spelling	.16	.14	.06	-.0807	.04	-.12	-.0965	-.13
Phonological awareness								
Pre-CTOPPP/CTOPP	.15	.16	.18	.0245	.14	†	†	-.09
Language								
PPVT	.23	.21	.15	-.0666	.01	.18	.0196	.06
TOLD	.42*	.38*	.17	-.2026	.04	.06	-.0558	-.07
Behavior								
SSRS Social Skills	.03	-.01	-.18	-.1683	-.18	†	†	-.05
SSRS Problem Behavior ⁴	-.39	-.34	-.14	.1997	.11	†	†	.46
PLBS/LBS	.12	.06	-.18	-.2367	-.26	†	†	-.32
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	.23	.26	.39	.1262	.19	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	-.43	-.36	-.07	.2828	.11	†	†	†
Arnett Harshness ⁵	-.34	-.35	-.38	-.0253	-.17	†	†	†
Arnett Permissiveness ⁵	1.28*	1.06*	.13	-.9123	-.04	†	†	†
Arnett Positive Interactions	.40	.40	.38	-.0121	.15	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Let's Begin with the Letter People: **University of Texas Health Science Center at Houston (Texas site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ESs) and slope effect sizes (ES_{slope}) are presented in table A-13.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 20 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means or in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment and (b) spring pre-kindergarten assessment, but there was a statistically reliable difference in the rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = .45; $ES_{\text{slope}} = .4783$, $p < .01$) favoring the treatment group. Finally, there were no statistically detectable differences on (d) the spring kindergarten assessment and (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three mathematics measures, we conclude that *Let's Begin with the Letter People* did not have an effect on mathematics development relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Language Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading

assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten. There was no statistically significant difference between groups on (d) the spring kindergarten assessment or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and TOLD Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted

in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the two language measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Problem Behaviors subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavior measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-13.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically significant differences between groups on (a) the means from the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant difference between groups (b) the spring pre-kindergarten observation ($ES_C = .82, p < .05$), such that treatment classrooms were rated as providing a more positive classroom environment. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Let's Begin with the Letter People* had no effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the following covariates: *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site.

On the Arnett Detachment scale, there were no statistically detectable differences between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there was no statistically detectable difference between groups on (a) the means from the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant on difference between groups at the time of (b) the spring pre-kindergarten observation ($ES_C = -.95, p < .05$), such that relative to control group teachers, treatment group teachers were rated as exhibiting less irritation toward the children and being less likely to use threats to manage children's behaviors. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there was a statistically reliable difference between groups in (a) the means from the fall pre-kindergarten observation ($ES_C = .99, p < .01$) and (c) the rate of change from the fall to spring observation (difference in rates of growth = $-.35$; $ES_{Slope} = -1.016, p < .05$). The statistically reliable difference in Arnett Permissiveness scores on the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the study curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the treatment group on the Arnett Harshness

($ES_c = 1.23, p < .01$). However, no statistically detectable differences were found between groups on (b) the spring pre-kindergarten observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.⁸

On the Arnett Positive Interactions scale, there were no statistically detectable differences between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Let's Begin with the Letter People*

The impact of *Let's Begin with the Letter People* on the child- and classroom-level measures is summarized in table A-13.

⁸ Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-13. Secondary analysis results for *Let's Begin with the Letter People*

Measure	RM analysis	RM analysis	RM analysis	Fall-Spring	ANCOVA ²	RM analysis	Spring Pre-K-K	ANCOVA	
	start of treatment ¹	Fall Pre-K	Spring Pre-K		slope	Spring Pre-K	kindergarten	slope	kindergarten
Student-level effect sizes									
Mathematics									
WJ Applied Problems	-.08	-.09	-.10	-.0179	-.03	-.13	-.0151	-.10	
CMA-A Mathematics Composite	.15	.15	.15	.0039	.12	-.07	-.1165	-.12	
Shape Composition ³	-.40*	-.28	.21	.4783**	.26	-.06	-.1427	-.00	
Reading									
TERA	-.03	-.03	.02	.0411	.04	-.13	-.0766	-.12	
WJ Letter Word Identification	-.16	-.11	.10	.2042	.19	-.18	-.1516	-.19	
WJ Spelling	-.03	.01	.17	.1565	.15	-.06	-.1239	-.13	
Phonological awareness									
Pre-CTOPPP/CTOPP	.12	.08	-.13	-.2006	-.16	†	†	-.13	
Language									
PPVT	.14	.11	-.03	-.1348	-.08	.00	.0133	-.02	
TOLD	.06	.07	.08	.0147	.08	-.12	-.1086	-.16	
Behavior									
SSRS Social Skills	-.46	-.43	-.27	.1520	.02	†	†	.24	
SSRS Problem Behavior ⁴	-.13	-.12	-.06	.0578	.02	†	†	.06	
PLBS/LBS	-.16	-.21	-.44	-.2154	-.35	†	†	-.10	
Classroom-level effect sizes									
Global classroom quality									
ECERS-R	-.07	.10	.82*	.7064	.74	†	†	†	
Teacher-child interaction									
Arnett Detachment ⁵	.22	.17	-.07	-.2316	-.15	†	†	†	
Arnett Harshness ⁵	.01	-.17	-.95*	-.7545	-.85	†	†	†	
Arnett Permissiveness ⁵	1.23**	.99**	-.05	-1.016*	-.29	†	†	†	
Arnett Positive Interactions	.02	.11	.48	.3693	.37	†	†	†	

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Early Literacy and Learning Model (ELLM): University of North Florida (Florida-UNF site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The University of North Florida (Florida-UNF) research team implemented its evaluation in three separate sites; table A-14 presents results for the combined analysis. Our discussion of the results focuses on the combined analysis of the three sites. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-14.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 28 days (including Saturdays, Sundays, and holidays) in County A, 27 days in County B, and 21 days in County C.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means or the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included

the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, and ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment and (b) the spring pre-kindergarten assessment, but there was (c) a statistically reliable difference between groups in the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = 8.54; $ES_{\text{slope}} = .3179$, $p < .01$), such that children in the treatment group learned at a faster rate than children in the control group. In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Spelling test. There were no statistically detectable differences (d) on the spring kindergarten assessment or (e) in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. There was a (d) statistically significant difference between groups on the spring kindergarten assessment ($ES_s = .34, p < .05$) favoring the treatment group, but there was no statistically detectable difference (e) in rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. There was a (d) statistically significant difference between groups on the spring kindergarten assessment ($ES_s = .44, p < .05$) favoring the treatment group, but there was no statistically detectable difference in (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten, but there was a significant difference in the covariate-adjusted means for spring kindergarten ($ES_s = .39, p < .01$).

Based on the analyses for the two language measures, we conclude that the *ELLM* did not have a statistically detectable effect on language development relative to the control condition in pre-kindergarten but had a positive effect in kindergarten.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measures, there were no statistically detectable differences between groups on (a) the means from the fall pre-kindergarten assessment, (b) the spring pre-kindergarten assessment, or (c) rate of change from the fall to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA,

there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{slope}) are presented in table A-14.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on the overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that the *ELLM* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for the *Early Literacy and Learning Model*

The impact of *Early Literacy and Learning Model* on the child- and classroom-level measures is summarized in table A-14.

Table A-14. Secondary analysis results for *Early Literacy and Learning Model*

Measure	RM analysis	RM analysis	RM analysis	Fall-Spring slope	ANCOVA ²	RM analysis	Spring Pre-K-	ANCOVA
	start of treatment ¹	Fall Pre-K	Spring Pre-K		Spring Pre-K	kindergarten	Spring K slope	kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.06	.06	.10	.0318	-.01	.26	.0885	.28
CMA-A Mathematics Composite	.01	.01	.01	.0001	.01	-.05	-.0330	-.09
Shape Composition ³	.17	.11	-.14	-.2432	-.19	.03	.0881	-.02
Reading								
TERA	.12	.13	.15	.0217	.04	.30	.0788	.20
WJ Letter Word Identification	-.14	-.12	-.05	.0713	-.04	.00	.0260	.03
WJ Spelling	-.29	-.21	.11	.3179**	.21	.04	-.0368	.08
Phonological awareness								
Pre-CTOPPP/CTOPP	-.06	-.01	.18	.1866	.14	†	†	.08
Language								
PPVT	.28	.26	.17	-.0843	-.06	.34*	.0891	.12
TOLD	-.01	.02	.15	.1328	.05	.44*	.1531	.39**
Behavior								
SSRS Social Skills	-.26	-.23	-.06	.1598	.08	†	†	.27
SSRS Problem Behavior ⁴	-.38*	-.35	-.24	.1073	-.01	†	†	.23
PLBS/LBS	.05	.07	.14	.0722	.07	†	†	.04
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	-.69	-.65	-.48	.1661	-.14	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	-.07	-.14	-.41	-.2668	-.02	†	†	†
Arnett Harshness ⁵	.08	-.01	-.40	-.3783	.02	†	†	†
Arnett Permissiveness ⁵	.09	.03	-.24	-.2666	-.10	†	†	†
Arnett Positive Interactions	-.53	-.38	.29	.6519	-.01	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Language-Focused Curriculum: University of Virginia (Virginia site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_s) and slope effect sizes (ES_{Slope}) are presented in table A-15.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 28 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, but there was (c) a statistically significant difference between groups in the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = 4.91; $ES_{Slope} = .2943$, $p < .05$). There were no statistically detectable differences (d) on the spring kindergarten assessment and (e) in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that the *Language-Focused Curriculum* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that the *Language-Focused Curriculum* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the phonological awareness measures, we conclude that the *Language-Focused Curriculum* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and TOLD Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the two language measures, we conclude that the *Language-Focused Curriculum* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten. There was no statistically significant difference in the covariate-adjusted means for the spring kindergarten assessment.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from

fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that the *Language-Focused Curriculum* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-15.

Overall classroom environment and teacher-child relationships

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year. We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year. We did not conduct analyses using the ECERS and Arnett data for this site because of data integrity concerns. During the baseline data collection, one observer completed the observational ratings in 8 of the 12 classrooms at this research site. It was later determined that the ECERS-R and Arnett ratings from these eight classrooms were inflated. Due to concerns with the integrity of the data from these eight classrooms, the decision was made to exclude the classroom quality and teacher-child relationships data for this site from the report.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBR] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Language-Focused Curriculum*

The impact of *Language-Focused Curriculum* on the child- and classroom-level measures is summarized in table A-15.

Table A-15. Secondary analysis results for *Language-Focused Curriculum*

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	-.18	-.11	.20	.2943*	.27	.11	-.0464	.12
CMA-A Mathematics Composite	-.02	.00	.08	.0824	.12	.00	-.0423	.02
Shape Composition ³	.08	.08	.08	.0055	-.01	.06	-.0126	.03
Reading								
TERA	.04	.07	.16	.0955	.09	.05	-.0617	-.07
WJ Letter Word Identification	.09	.09	.11	.0228	.11	.02	-.0509	-.05
WJ Spelling	.07	.10	.25	.1416	.29	.11	-.0722	.06
Phonological awareness								
Pre-CTOPPP/CTOPP	-.12	-.06	.20	.2565	.28	†	†	.03
Language								
PPVT	-.16	-.12	.02	.1423	.12	-.09	-.0612	-.03
TOLD	-.12	-.10	.01	.1019	-.04	-.07	-.0393	-.03
Behavior								
SSRS Social Skills	.20	.08	-.42	-.4904	-.51	†	†	-.07
SSRS Problem Behavior ⁴	.40	.39	.37	-.0173	.21	†	†	-.05
PLBS/LBS	-.28	-.28	-.27	.0096	-.25	†	†	.10
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	—	—	—	—	—	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	—	—	—	—	—	†	†	†
Arnett Harshness ⁵	—	—	—	—	—	†	†	†
Arnett Permissiveness ⁵	—	—	—	—	—	†	†	†
Arnett Positive Interactions	—	—	—	—	—	†	†	†

— Not available. Data were collected but not reported.

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

²The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³Building Blocks, Shape Composition task

⁴Higher scores on this scale represent more negative child behaviors.

⁵Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

For the *Language-Focused Curriculum*, we did not conduct analyses using the ECERS and Arnett data because of unreliable data. During the baseline data collection, one observer completed the observational ratings in eight of the 12 classrooms at this research site. It was later determined that the ECERS-R and Arnett ratings from these eight classrooms were inflated. Due to concerns with the integrity of the data from these eight classrooms, the decision was made to exclude the classroom quality and teacher-child relationships data for this site from the report. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Literacy Express: **Florida State University (Florida-FSU site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-16.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 42 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that *Literacy Express* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's

education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment and (b) spring pre-kindergarten assessment, but there was significant difference between groups in (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = 2.10; $ES_{\text{slope}} = .2815, p < .05$). There were no statistically detectable differences on the (d) spring kindergarten assessment and (e) no statistically detectable difference between groups in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there was (a) a significant difference between groups on the fall assessment ($ES_s = .44, p < .05$). This result could indicate either an early treatment effect or failure of random assignment to produce equivalent groups. We extrapolated back to the start of the school year and found a statistically reliable difference in means ($ES_s = .47, p < .05$) at the start of the year. This difference suggests, but does not prove, nonequivalence of treatment and control groups. Because there was no evidence on any other measure of nonequivalence between groups at the start of treatment, we considered the groups to be equivalent. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

In the spring of pre-kindergarten, there were no statistically detectable differences between groups on (b) the covariate-adjusted means at the spring pre-kindergarten assessment and (c) the rates of growth fall to spring of the pre-kindergarten year. There were no statistically detectable differences on the (d) spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *Literacy Express* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups in the (a) fall pre-kindergarten assessment, or (b) spring pre-kindergarten assessment. There was, however, a statistically

significant difference in (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (differences in rates of growth = 1.35; $ES_{\text{slope}} = .3217, p < .05$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Literacy Express* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and TOLD Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the two language measures, we conclude that *Literacy Express* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or on (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on (a) the means from the fall pre-kindergarten, (b) the spring pre-kindergarten assessment, or the (c) rate of change from the fall to spring assessment. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we

could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was a significant difference in the covariate-adjusted means for spring kindergarten ($ESs = -.38, p < .05$), such that children in the treatment group showed weaker learning behaviors than children in the control group.

Based on the analyses of the behavioral measures, we conclude that *Literacy Express* did not have a statistically detectable effect on social and learning behaviors relative to the control condition at spring of pre-kindergarten, but did have a negative effect on social and learning behaviors relative to the control condition at spring of kindergarten.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-16.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. There was a statistically significant difference between groups on the (b) spring pre-kindergarten observation ($ES_C = 1.29, p < .05$). On the ANCOVA, a statistically significant difference was obtained on the spring pre-kindergarten observation ($ES_C = 1.22, p < .05$) favoring the treatment group.

Based on analyses of the ECERS-R, we conclude that *Literacy Express* had a positive effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Literacy Express* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Literacy Express*

The impact of *Literacy Express* on the child- and classroom-level measures is summarized in table A-16.

Table A-16. Secondary analysis results for *Literacy Express*

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K- Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.02	.03	.05	.0260	-.04	-.02	-.0379	-.09
CMA-A Mathematics Composite	-.17	-.15	-.02	.1192	.03	-.21	-.0975	-.17
Shape Composition ³	-.15	-.12	-.01	.1057	.05	-.14	-.0664	-.07
Reading								
TERA	-.19	-.12	.17	.2815*	.23	-.11	-.1518	-.02
WJ Letter Word Identification	.47*	.44*	.30	-.1341	-.02	.08	-.1154	-.29
WJ Spelling	.14	.12	.05	-.0671	-.07	.06	.0037	.06
Phonological awareness								
Pre-CTOPPP/CTOPP	-.26	-.19	.14	.3217*	.15	†	†	.08
Language								
PPVT	.05	.07	.17	.0963	.06	.16	-.0041	.09
TOLD	-.25	-.21	-.04	.1687	.04	.10	.0743	.13
Behavior								
SSRS Social Skills	.44	.35	-.06	-.4022	-.21	†	†	-.37
SSRS Problem Behavior ⁴	-.60*	-.54	-.31	.2279	.00	†	†	.22
PLBS/LBS	.42	.38	.17	-.2053	-.02	†	†	-.38*
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	-.14	.12	1.29*	1.1353	1.22*	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	.46	.17	-1.09	-1.230	-1.17	†	†	†
Arnett Harshness ⁵	.03	-.13	-.84	-.6959	-.94	†	†	†
Arnett Permissiveness ⁵	-1.12	-.82	.51	1.2987	.61	†	†	†
Arnett Positive Interactions	-.89	-.62	.56	1.1518	1.04	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

DLM Early Childhood Express supplemented with Open Court Reading Pre-K: **Florida State University (Florida-FSU site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_s) and slope effect sizes (ES_{Slope}) are presented in table A-17.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 28 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment or (b) the rates of growth from fall pre-kindergarten to spring pre-kindergarten, but there was a statistically significant difference on the (c) spring pre-kindergarten assessment ($ES_s = .36, p < .01$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Applied Problems. There was a statistically significant difference on the (d) spring kindergarten assessment ($ES_s = .48, p < .001$), but no statistically significant difference on (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, but there was a statistically significant difference between groups in the covariate-adjusted spring kindergarten means ($ES_s = .31, p < .05$). We conclude there was no effect of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* on the WJ Applied Problems for the spring pre-kindergarten assessment; however, *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had a positive effect on the WJ Applied Problems for the spring of kindergarten assessment.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences

between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Across the three math measures, we did not obtain a consistent pattern of results and concluded that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have an impact on the mathematics outcome relative to the control condition. Although we did not obtain consistent evidence across the mathematics measures, the analyses of the WJ Applied Problems are consistent with the pattern of results from the reading and phonological awareness data. We examined the intercorrelations among measures for the combined control group (i.e., children in the control condition across all research/grantee sites) for the fall assessment. Scores on the WJ Applied Problems were moderately correlated with each of the reading, phonological awareness, and language measures (TERA: $r = .60$; WJ Letter Word Identification: $r = .49$, $p < .0001$; WJ Spelling: $r = .46$, $p < .0001$; Pre-CTOPPP: $r = .48$, $p < .0001$; PPVT: $r = .63$, $p < .0001$; TOLD Grammatic Understanding subtest: $r = .47$, $p < .0001$).

Based on the analyses of the three mathematics measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there was no statistically detectable difference between groups on the (a) fall assessment, but there were statistically reliable differences on the (b) spring pre-kindergarten assessment ($ES_s = .68$, $p < .001$) and (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = 3.22; $ES_{\text{slope}} = .4052$, $p < .001$). Taken together these three results provide statistical evidence of a treatment effect of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* on the TERA. On the ANCOVA, there was a statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means ($ES_s = .40$, $p < .05$).

In addition, on the repeated measures analysis of the TERA, there was a statistically reliable difference between groups on the (d) spring kindergarten assessment ($ES_s = .76$, $p < .01$); there were no differences in the (e) rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were differences between groups on the (a) fall assessment ($ES_s = .41$, $p < .05$) and (b) spring pre-kindergarten assessment ($ES_s = .51$, $p < .01$), but not in the (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten. The statistically reliable difference in WJ Letter Word Identification scores at baseline suggests either the nonequivalence of treatment and control groups or an early treatment effect. To examine the possibility of an early treatment effect, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* group on the WJ Letter Word Identification test ($ES_s = .39$, $p < .05$). For the following reasons, this difference suggests, but does not prove, nonequivalence of treatment and control groups. The extrapolation is based on the average score for each group at the fall assessment and the rate of growth from fall to spring for each group. Although the slope from fall to spring is our best estimate of the rate of growth from the beginning of the school year to the fall assessment, if there were an early treatment effect, there is no theoretical reason to assume that growth was constant from the start of the school year to the spring assessment. The WJ Letter Word Identification test has a high floor—that is, there are few items for younger children and one additional correct answer can result in substantial differences in

the standardized scores. For example, for a child age 4 years and 6 months, three correct answers yield a standardized score of 89 and four correct answers a score of 95. The fall assessment average standardized raw score for the control group was 91.53; their spring standardized raw score was 95.60.⁹ The difference between the fall and spring assessments is less than one correct answer. Because the initial items on the test are letter identification items, curricula that focus first on learning letters have the possibility to produce a rapid early treatment effect.

In the spring of pre-kindergarten, statistically reliable differences were obtained on the WJ Letter Word Identification test ($ES_s = .51, p < .01$); however there was no statistically detectable difference in slopes from fall to spring of the pre-kindergarten year. Consequently the difference obtained at spring could simply reflect the difference obtained at baseline. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the spring kindergarten assessment, statistically reliable differences were obtained on the WJ Letter Word Identification test ($ES_s = .50, p < .01$); however, there was no statistically detectable difference in rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment or in (b) the rates of growth from fall pre-kindergarten to spring pre-kindergarten. However, there was a statistically reliable difference on the (c) spring pre-kindergarten assessment ($ES_s = .46, p < .01$). Taken together, these results do not meet the criteria for establishing statistical evidence of a treatment effect for a single outcome measure in the pre-kindergarten year. On the ANCOVA, there was a statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means ($ES_s = .27, p < .05$).

On the repeated measures analysis of the WJ Spelling, there were no statistically significant group differences on the (d) spring kindergarten assessment, or in the (e) rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had a positive effect on reading relative to the control condition in pre-kindergarten.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) pre-kindergarten fall assessment, or the (b) the rate of growth from fall pre-kindergarten to spring pre-kindergarten, but a statistically reliable difference between groups was found at the spring pre-kindergarten assessment ($ES_s = .32, p < .05$). On the ANCOVA for the Pre-CTOPPP, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means.

⁹ Schrank, F.A. and Woodcock, R.W. (2001). WJ III Compuscore and Profiles Program [Computer software]. *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

For the ANCOVA on the kindergarten CTOPP data, there was a statistically significant difference between groups in the covariate-adjusted spring kindergarten means ($ES_s = .38, p < .05$).

Based on the analyses of the phonological awareness measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had no effect on phonological awareness relative to the control condition in pre-kindergarten and a positive effect in kindergarten.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there was no statistically detectable difference between groups on (a) the fall assessment. In the spring of the pre-kindergarten year, there was a statistically reliable mean difference between groups ($ES_s = .40, p < .05$); however, there was (c) no statistically reliable difference between groups in the slopes. On the ANCOVA for the PPVT, there were no statistically significant differences in covariate-adjusted means for the spring pre-kindergarten assessment.

In spring of the kindergarten year, there was a statistically reliable mean difference between groups on the PPVT ($ES_s = .48, p < .01$); there was no statistically significant difference in the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA analysis for the PPVT, there was no statistically significant difference in covariate-adjusted means for the spring kindergarten assessment.

For the TOLD Grammatical Understanding subtest, there was (a) a statistically reliable difference in means favoring children in the *Open Court Reading Pre-K* condition on the fall assessment ($ES_s = .38, p < .05$). We extrapolated scores back to the beginning of the year and found no statistically detectable differences at the start of the school year between the two groups. This result suggests that differences at the fall assessment reflect early treatment effects. In the spring of the pre-kindergarten year, there was (b) a statistically reliable mean difference in scores on the TOLD Grammatical Understanding subtest ($ES_s = .40, p < .01$), but there was (c) no statistically detectable difference between groups in the rates of growth from fall to spring pre-kindergarten. On the ANCOVA for the TOLD Grammatical Understanding subtest, there were no statistically detectable differences in covariate-adjusted means for the spring pre-kindergarten assessment.

In spring of the kindergarten year, there was a statistically reliable difference in means on the TOLD favoring children in the *Open Court* condition ($ES_s = .46, p < .01$). There was no statistically detectable difference between groups in the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA analyses for TOLD Grammatical Understanding, there was no statistically detectable difference between groups in covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the two language measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have an effect on language development relative to the control condition at spring of pre-kindergarten. However, *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had a positive effect on language development relative to the control condition by spring of kindergarten.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS

Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, or (b) spring pre-kindergarten assessment, but a statistically reliable difference between groups was found on (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = -6.4, $ES_{\text{slope}} = -.4253$, $p < .05$). On the ANCOVA, there were no differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, or (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the four behavioral measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{slope}) are presented in table A-17.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on overall classroom environment relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition,

for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, there was no statistically detectable difference between groups on the spring pre-kindergarten.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

The impact of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* on the child- and classroom-level measures is summarized in table A-17.

Table A-17. Secondary analysis results for DLM Early Childhood Express supplemented with Open Court Reading Pre-K

Measure	RM analysis	RM Analysis	RM analysis	Fall-Spring	ANCOVA ²	RM analysis	Spring Pre-K	ANCOVA
	start of treatment ¹	Fall Pre-K	Spring Pre-K	slope	Spring Pre-K	kindergarten	Spring K slope	kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.17	.21	.36**	.1497	.19	.48***	.0608	.31*
CMA-A Mathematics Composite	-.14	-.08	.17	.2526	.22	.13	-.0232	.10
Shape Composition ³	-.16	-.08	.24	.3154	.21	.09	-.0806	.10
Reading								
TERA	.17	.27	.68***	.4052***	.40*	.76**	.0418	.33
WJ Letter Word Identification	.39*	.41*	.51**	.0945	.23	.50**	-.0070	.13
WJ Spelling	.21	.26	.46**	.1984	.27*	.22	-.1282	.14
Phonological awareness								
Pre-CTOPPP/CTOPPP	.02	.08	.32*	.2377	.25	†	†	.38*
Language								
PPVT	.31	.33	.40*	.0754	.15	.48**	.0408	.22
TOLD	.38	.38*	.40**	.0207	.26	.46**	.0329	.27
Behavior								
SSRS Social Skills	.42	.32	-.11	-.4253*	-.28	†	†	-.18
SSRS Problem Behavior ⁴	-.04	-.01	.11	.1247	.11	†	†	.01
PLBS/LBS	-.08	-.09	-.16	-.0671	-.08	†	†	-.13
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	-.21	-.11	.34	.4320	.55	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	.06	.04	-.06	-.0928	-.22	†	†	†
Arnett Harshness ⁵	-.37	-.43	-.70	-.2574	-.85	†	†	†
Arnett Permissiveness ⁵	-.37	-.29	.05	.3327	.23	†	†	†
Arnett Positive Interactions	-.66	-.46	.43	.8638	.53	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

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

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Pre-K Mathematics supplemented with DLM Early Childhood Express Math software:

University of California, Berkeley and University at Buffalo, State University of New York (California and New York sites)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-18.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 19 days (including Saturdays, Sundays, and holidays) in California and 14 days in New York.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there was no statistically detectable difference between groups on the (a) fall assessment, but there were (b) statistically reliable differences between groups on means for the spring pre-kindergarten assessment ($ES_S = .44, p < .01$), and (c) in the rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = .08; $ES_{Slope} = .3632, p < .01$). Taken together, these three results provide clear evidence of an effect. There was (d) no difference in means for the spring kindergarten assessment, but there was (e) a statistically reliable difference in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rates of growth = -.04; $ES_{Slope} = -.1690, p < .01$). These last two results indicate that from spring of pre-kindergarten through spring of kindergarten, children who had been in *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* were learning at a *slower* rate relative to children who had been in the pre-kindergarten control classrooms. By the spring kindergarten assessment, there was no statistically detectable difference between the two groups. On the ANCOVA, there was a statistically reliable difference between groups in the covariate-adjusted spring pre-kindergarten means ($ES_S = .35, p < .01$), and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were statistically reliable differences between groups on: (a) the means from the fall assessment ($ES_S = .25, p < .05$), (b) the means from spring pre-kindergarten assessment ($ES_S = .96, p < .001$), and (c) the rates of growth from fall pre-kindergarten to spring pre-kindergarten

(difference in rates of growth = .61; $ES_{\text{slope}} = .6999$, $p < .0001$). The difference between groups at the fall assessment could reflect a failure of randomization to produce equivalent groups or an early treatment effect. We extrapolated back to the start of the school year; there was no statistically significant difference in means at the start of the school year. This finding suggests that the groups were equivalent at the beginning of the year and the observed difference in the fall reflected an early treatment effect. Taken together, these results provide evidence of a positive effect of *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* on Shape Composition relative to the control condition.

There was a statistically reliable difference between groups on (d) the means at the spring kindergarten assessment ($ES_s = p < .001$), and (e) the rates of growth from spring pre-kindergarten to spring kindergarten (difference in rates of growth = -.26; $ES_{\text{slope}} = -.2986$, $p < .0001$). These last two results indicate that from spring of pre-kindergarten through spring of kindergarten, children who had been in the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* were learning at a *slower* rate relative to children who had been in the pre-kindergarten control classrooms. Despite this slower rate of growth, the advantage obtained by the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* children by the spring of pre-kindergarten remained through spring of the kindergarten year. On the ANCOVA, there were statistically reliable differences between groups in the covariate-adjusted spring pre-kindergarten means ($ES_s = .91$, $p < .0001$), and in the covariate-adjusted spring kindergarten means ($ES_s = .30$, $p < .01$).

Based on the analyses for the three mathematics measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* had a positive effect on mathematics at the end of pre-kindergarten relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Testing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on these analyses for the phonological awareness measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the two language measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, or (b) spring pre-kindergarten assessment, but there was a statistically reliable difference between groups in the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = 4.38, $ES_{\text{slope}} = .3040$, $p < .05$). On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{slope}) are presented in table A-18.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically significant difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the teacher-child relationship measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

The impact of *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* on the child- and classroom-level measures is summarized in table A-18.

Table A-18. Secondary analysis results for Pre-K Mathematics supplemented with DLM Early Childhood Express Math software

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.19	.19	.22	.0229	.16	.13	-.0437	.05
CMA-A Mathematics Composite	-.02	.07	.44**	.3632**	.35**	.13	-.1690**	.10
Shape Composition ³	.08	.25*	.96***	.6999****	.91****	.41***	-.2986****	.30**
Reading								
TERA	.14	.14	.13	-.0099	.00	.31	.0969	.08
WJ Letter Word Identification	-.18	-.15	-.01	.1349	.06	.22	.1233	.21
WJ Spelling	.14	.15	.20	.0445	.17	.03	-.0881	-.08
Phonological awareness								
Pre-CTOPPP/CTOPP	-.17	-.13	.04	.1663	.11	†	†	-.11
Language								
PPVT	.02	.05	.17	.1223	.18	.11	-.0363	.09
TOLD	.12	.13	.17	.0385	.07	.08	-.0473	-.03
Behavior								
SSRS Social Skills	-.17	-.10	.22	.3040*	.24	†	†	.06
SSRS Problem Behavior ⁴	.10	.07	-.09	-.1523	-.11	†	†	-.01
PLBS/LBS	-.10	-.06	.09	.1494	.08	†	†	.01
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	-.58	-.46	.05	.5040	-.22	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	.12	.02	-.37	-.3841	.09	†	†	†
Arnett Harshness ⁵	.16	.16	.18	.0174	.31	†	†	†
Arnett Permissiveness ⁵	.12	.02	-.45	-.4521	-.36	†	†	†
Arnett Positive Interactions	-.27	-.19	.16	.3471	-.24	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$; *** $p < .001$; **** $p < .0001$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Project Approach: Purdue University and University of Wisconsin-Milwaukee (Wisconsin site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-19.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 13 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, however, there was a statistically significant difference in the (e) rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = 2.37; $ES_{Slope} = .1043$, $p < .05$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment or (b) spring pre-kindergarten assessment; however, there was a statistically reliable difference in the (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = .06; $ES_{Slope} = .2640$, $p < .05$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Applied Problems. There were no statistically detectable differences between groups on (d) the spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no detectable significant differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment or the (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten. There were no statistically detectable differences between groups on the (d) spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that *Project Approach* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were (a) no statistically detectable differences between groups on the (a) fall assessment or (b) spring pre-kindergarten assessment; however, there was a statistically reliable difference between groups on the (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = -2.1; $ES_{Slope} = -.2202$, $p < .05$). We do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the TERA. There was no statistically detectable difference between groups on (d) the spring kindergarten assessment or the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically significant differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, and (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. There was no statistically significant difference on the (d) spring kindergarten assessment; however, there was a statistically significant difference between groups in the (e) rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = -5.7; $ES_{Slope} = -.2102$, $p < .01$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, but there was a statistically significant difference between groups in the covariate-adjusted spring kindergarten means ($ES_s = -.44$, $p < .05$).

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Project Approach* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the

covariate-adjusted spring pre-kindergarten means. For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the phonological awareness measures, we conclude that *Project Approach* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment and (b) the spring pre-kindergarten assessment, or (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten. There were no statistically detectable differences between groups on (d) the spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten, the spring kindergarten assessment, or the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

Based on the analyses for the two language measures, we conclude that *Project Approach* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social System Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten, but there was a significant difference between groups at spring kindergarten ($ES_s = -.44, p < .05$).

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no

statistically detectable differences in the covariate-adjusted means for the spring pre-kindergarten assessment, but there was a significant difference between groups at spring kindergarten ($ES_s = .49, p < .05$).

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment or (b) spring pre-kindergarten assessment. There was a statistically significant difference in the rate of growth from fall to spring pre-kindergarten (different in rates of growth = -2.9, $ES_{Slope} = -.2922, p < .05$). On the ANCOVA, there was a statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment ($ES_s = -.37, p < .05$).

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Project Approach* did not have a statistically detectable effect on social and learning behaviors relative to the control condition during pre-kindergarten, but *Project Approach* had a negative effect on behavior by spring of the kindergarten year relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{Slope}) are presented in table A-19.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Project Approach* did not have a statistically detectable on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change

from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, there was no statistically detectable difference on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the four teacher-child relationship measures, we conclude that *Project Approach* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Project Approach*

The impact of *Project Approach* on the child- and classroom-level measures is summarized in table A-19.

Table A-19. Secondary analysis results for Project Approach

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K- Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	-.08	-.05	.07	.1201	-.11	.27	.1043*	.02
CMA-A Mathematics Composite	-.15	-.09	.18	.2640*	.00	.22	.0240	-.03
Shape Composition ³	-.04	.02	.27	.2478	-.13	.24	-.0173	-.11
Reading								
TERA	.42	.36	.14	-.2202*	-.23	.29	.0807	-.18
WJ Letter Word Identification	.49	.47	.42	-.0548	-.05	.03	-.2102**	-.44*
WJ Spelling	.34	.33	.27	-.0561	-.19	.14	-.0682	-.35
Phonological awareness								
Pre-CTOPPP/CTOPP	.15	.13	.05	-.0752	-.27	†	†	-.17
Language								
PPVT	-.06	-.02	.16	.1775	.07	.10	-.0352	-.10
TOLD	-.07	-.03	.15	.1752	-.08	.32	.0875	.04
Behavior								
SSRS Social Skills	-.07	-.05	.04	.0891	.06	†	†	-.44*
SSRS Problem Behavior ⁴	.34	.37	.50	.1212	.23	†	†	.49*
PLBS/LBS	.06	-.01	-.31	-.2922*	-.37*	†	†	-.42
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	1.07	.84	-.19	-1.0030	-.27	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	.12	.20	.57	.3596	.37	†	†	†
Arnett Harshness ⁵	1.10	1.06	.86	-.1903	-.21	†	†	†
Arnett Permissiveness ⁵	-.28	-.30	-.43	-.1191	-.57	†	†	†
Arnett Positive Interactions	-.98	-.98	-.99	-.0091	.69	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$; ** $p < .01$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Project Construct: **University of Missouri-Columbia (Missouri site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_s) and slope effect sizes (ES_{slope}) are presented in table A-20.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 42 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on (a) the fall assessment, or (c) in the rate of growth from fall pre-kindergarten to spring pre-kindergarten, but there was (b) a significant difference between groups on the spring pre-kindergarten assessment ($ES_s = -.42$, $p < .01$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the Shape Composition task. There was no statistically detectable difference on the spring kindergarten assessment, but there was a statistically significant difference between groups in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = .01; $ES_s = .2846$, $p < .001$). The observed difference in rate of growth from spring pre-kindergarten to spring kindergarten with no statistically significant difference in means on the spring pre-kindergarten and spring kindergarten assessments provides inconclusive evidence that children from *Project Construct* were learning at a slower rate in the year after the pre-kindergarten intervention. On the ANCOVA, there was a statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means ($ES_s = -.44$, $p < .05$), and no statistically significant differences between groups in the covariate-adjusted spring kindergarten means.

The ANCOVA analysis indicates that from the fall assessment to the spring pre-kindergarten assessment students did not gain as much relative to students in the control classrooms.

Based on the analyses for the three mathematics measures, we conclude that *Project Construct* did not have a statistically detectable effect on the mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Project Construct* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Project Construct* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT], and TOLD Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

Based on the analyses for the two language measures, we conclude that *Project Construct* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no

statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Project Construct* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{slope}) are presented in table A-20.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Project Construct* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Project Construct* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Project Construct*

The impact of *Project Construct* on the child- and classroom-level measures is summarized in table A-20.

Table A-20. Secondary analysis results for *Project Construct*

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K- Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	-.07	-.04	.06	.0990	.10	.08	.0107	.07
CMA-A Mathematics Composite	-.02	-.04	-.11	-.0725	-.07	-.06	.0287	-.08
Shape Composition ³	-.10	-.16	-.42**	-.2492	-.41*	.12	.2846***	.13
Reading								
TERA	-.07	-.06	.00	.0576	.18	-.03	-.0166	.12
WJ Letter Word Identification	.00	-.01	-.05	-.0370	-.07	.16	.1129	.22
WJ Spelling	-.31	-.28	-.15	.1268	-.02	.00	.0798	.13
Phonological awareness								
Pre-CTOPPP/CTOPP	.37	.32	.10	-.2176	-.07	†	†	-.12
Language								
PPVT	.02	.02	.03	.0129	-.03	.10	.0339	.11
TOLD	.23	.18	-.05	-.2196	-.13	.01	.0305	-.08
Behavior								
SSRS Social Skills	.22	.22	.22	.0005	.07	†	†	.12
SSRS Problem Behavior ⁴	-.22	-.19	-.08	.1060	.07	†	†	.07
PLBS/LBS	.20	.16	.00	-.1585	-.11	†	†	-.02
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	.27	.32	.54	.2112	.35	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	.27	.24	.12	-.1183	-.24	†	†	†
Arnett Harshness ⁵	-.95	-.80	-.13	.6485	-.02	†	†	†
Arnett Permissiveness ⁵	.42	.34	-.02	-.3532	-.18	†	†	†
Arnett Positive Interactions	.40	.41	.46	.0464	.39	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

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

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Ready, Set, Leap! **University of California, Berkeley (New Jersey site)**

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-21.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 35 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were (a) no statistically detectable differences between groups on the fall assessment, (b) a statistically reliable difference on the spring pre-kindergarten assessment ($ES_S = -.24$, $p < .05$), and (c) no statistically detectable difference in rate of growth from fall pre-kindergarten to spring pre-kindergarten. Taken together, these three results do not provide conclusive evidence of an effect of *Ready, Set, Leap!* relative to the control condition on the CMA-A Composite Score for the pre-kindergarten year. In addition, there were no statistically detectable differences between groups on (d) the spring kindergarten assessment or (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there was (a) a statistically reliable difference between groups on the fall assessment ($ES_S = .25$, $p < .05$), but none on the (b) spring pre-kindergarten assessment or (c) the rates of growth from fall pre-kindergarten to spring pre-kindergarten. The difference between groups on the fall assessment could reflect either the failure of randomization to produce equivalent groups or an early treatment effect (which was not sustained to the spring pre-kindergarten assessment). To examine the possibility of an early treatment effect, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the *Ready, Set, Leap!* group on the Shape Composition measure ($ES_S = .29$, $p < .05$). There was no statistically significant difference on the start of treatment extrapolated means on any of the other mathematics measures or other child measures. Given the lack of consistent

results across measures and because the test of the extrapolated means does not provide conclusive evidence of nonequivalence at the start of the treatment,¹⁰ we conclude the evidence is inconclusive for determining nonequivalence at the start of treatment.

In addition, on the Shape Composition task, there were no statistically significant differences between groups on (d) the spring kindergarten assessment or (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically significant differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on reading relative to the control condition.

¹⁰ For the following reasons, this difference suggests, but does not prove, nonequivalence of treatment and control groups. The extrapolation is based on the average score for each group for the fall assessment and the rate of growth from fall to spring for each group. Although the slope from fall to spring is our best estimate of the rate of growth from the beginning of the school year to the fall assessment, if there were an early treatment effect, there is no theoretical reason to assume that growth was constant from the start of the school year to the spring assessment. Across the nine academic outcomes, the Shape Composition was the only one on which there were statistically reliable differences for the fall assessment means or the start of treatment extrapolated means.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the phonological awareness measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT], Test of Language Development [TOLD] Grammatical Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatical Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the two language measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills, SSRS Problem Behaviors, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_c) and slope effect sizes (ES_{slope}) are presented in table A-21.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*,

teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained for the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained during the spring of pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Ready, Set, Leap!*

The impact of *Ready, Set, Leap!* on the child- and classroom-level measures is summarized in table A-21.

Table A-21. Secondary analysis results for *Ready, Set, Leap!*

Measure	RM analysis start of treatment ¹	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA ² Spring Pre-K	RM analysis kindergarten	Spring Pre-K Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.14	.12	.04	-.0753	.03	.00	-.0221	-.05
CMA-A Mathematics Composite	-.07	-.10	-.24*	-.1317	-.23	-.10	.0719	-.11
Shape Composition ³	.29*	.25*	.08	-.1665	-.01	.03	-.0277	-.06
Reading								
TERA	.06	.07	.08	.0125	.07	.01	-.0361	-.05
WJ Letter Word Identification	-.03	-.02	.01	.0360	.03	-.12	-.0728	-.06
WJ Spelling	.01	.04	.20	.1477	.13	.04	-.0829	.02
Phonological awareness								
Pre-CTOPPP/CTOPP	.03	.01	-.09	-.0913	-.06	†	†	-.02
Language								
PPVT	.19	.18	.15	-.0320	.01	-.02	-.0888	-.13
TOLD	.07	.04	-.11	-.1425	-.14	-.03	.0441	-.04
Behavior								
SSRS Social Skills	-.01	-.02	-.05	-.0317	-.05	†	†	-.03
SSRS Problem Behavior ⁴	.02	.01	-.03	-.0373	.00	†	†	.07
PLBS/LBS	.10	.09	.07	-.0279	.01	†	†	-.01
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	-.03	.01	.16	.1473	.32	†	†	†
Teacher-child interaction								
Arnett Detachment ⁵	.19	.19	.19	.0028	.17	†	†	†
Arnett Harshness ⁵	.23	.25	.30	.0561	.26	†	†	†
Arnett Permissiveness ⁵	-.36	-.34	-.24	.0903	-.09	†	†	†
Arnett Positive Interactions	-.08	-.05	.04	.0936	.15	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.

* $p < .05$

¹ The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

² The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

³ Building Blocks, Shape Composition task

⁴ Higher scores on this scale represent more negative child behaviors.

⁵ Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures

ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

**Appendix B:
Data Analysis Approach and Statistical Model**

Page intentionally left blank.

Research Questions

The Preschool Curriculum Evaluation Research (PCER) initiative focused on the impact of the intervention curricula on students' reading, phonological awareness, early language, early mathematics knowledge, and behavior (including social skills) at the end of pre-kindergarten and kindergarten. These domains of knowledge and skills are predictive of academic success in the early years of elementary school (Downer and Pianta 2006; Miles and Stipek 2006).

In addition, the PCER evaluation study also examined the impact of the curriculum interventions on teachers' classroom instructional practice, teacher-child interaction, and global classroom quality. These dimensions of early childhood programs have been posited as mediators (e.g., instructional practice) and moderators (e.g., teacher-child interaction, classroom quality) of the relation between early childhood curricula and child outcomes (Arnett 1989; Cost, Quality, and Child Outcomes Study 1995; Peisner-Feinberg and Burchinal 1997; Ruopp et al. 1979).

In sum, the research questions for the evaluation primarily concern student academic and behavioral outcomes and also include classroom outcomes due to their potentially mediating or moderating roles. The research questions are:

1. What is the impact of each of the 14 preschool curricula on preschool students' reading skills, phonological awareness, language development, mathematical knowledge, and behavior?
2. What is the impact of each of the 14 preschool curricula on these outcomes for students at the end of kindergarten?
3. What is the impact of each of the 14 preschool curricula on preschool classroom quality, teacher-child interactions, and instructional practices?

Data Collection, Sample, and Assignment

Data were collected in the fall (baseline) and spring of the pre-kindergarten year and in the spring of the kindergarten year to answer the research questions outlined above. Each research team recruited preschool programs, teachers, children, and parents for participation in the PCER evaluation study. Overall, 2,911 children, 315 preschool classrooms, and 208 preschools were involved in the PCER initiative.

The research teams recruited samples of convenience from local preschool programs willing to agree to the random assignment of classrooms or schools to treatment and control conditions. Table B-1 details the research teams, the curricula they evaluated, and the sample size and unit of assignment they used.

Variables Used in Analysis and General Data Issues

Time Structure of Data

Data were collected at three time points: fall pre-kindergarten (i.e., baseline), spring pre-kindergarten, and spring kindergarten. A common measurement battery was used to collect child, parent, teacher, and classroom observation (pre-kindergarten only) data at all research sites. For a few child-outcome measures, the assessment instruments changed between pre-kindergarten and kindergarten or changes in standardizations. Instruments included the Social Skills Rating System (SSRS) Socials and Problem Behaviors scales; the Preschool Learning Behaviors Scale (PLBS) (pre-kindergarten); the Learning Behaviors Scale (LBS) (kindergarten); the Preschool Comprehensive Test of Phonological and Print Processing, Elision subtest (Pre-CTOPPP) (pre-kindergarten); and the Comprehensive Test of Phonological Processing (CTOPP), Elision subtest. Classrooms were only observed during the intervention (pre-kindergarten) year, as the children all

Table B-1. Units of random assignment for evaluation of each curriculum

Research team (site)	Curricula	Treatment sample	Control sample	Students
Vanderbilt University (TN)	<i>Bright Beginnings</i>	7 classrooms	7 classrooms	T: 103
	<i>Creative Curriculum</i>	7 classrooms		C: 105
UNC-Charlotte (NC, GA)	<i>Creative Curriculum</i>	9 classrooms	9 classrooms	T: 97 C: 97
University of New Hampshire (NH)	<i>Creative Curriculum with Ladders to Literacy</i>	7 classrooms	7 classrooms	T: 62 C: 61
Success for All Foundation (NJ, KS, FL)	<i>Curiosity Corner</i>	10 Pre-K programs	8 Pre-K programs	T: 105 C: 110
University of Texas Health Science Center at Houston (TX)	<i>Doors to Discovery</i>	14 classrooms	15 classrooms	T: 101
	<i>Let's Begin with the Letter People</i>	15 classrooms		C: 96
University of North Florida (FL)	<i>Early Literacy and Learning Model</i>	14 classrooms ¹	14 classrooms ¹	T: 137 C: 107
University of Virginia (VA)	<i>Language-Focused Curriculum</i>	7 classrooms	7 classrooms	T: 97 C: 98
Florida State University (FL)	<i>DLM Early Childhood Express with Open Court Reading Pre-K</i>	5 Pre-K programs	6 Pre-K programs	T: 101
	<i>Literacy Express</i>	6 Pre-K programs		C: 97 T: 99
UC-Berkeley and University at Buffalo, SUNY (CA, NY)	<i>Pre-K Mathematics with DLM Early Childhood Express Math software</i>	20 classrooms	20 classrooms	T: 159 C: 157
Purdue University and University of WI-Milwaukee (WI)	<i>Project Approach</i>	7 classrooms	6 classrooms	T: 114 C: 90
University of Missouri-Columbia (MO)	<i>Project Construct</i>	10 Pre-K programs ¹	11 Pre-K programs ¹	T: 123 C: 108
UC-Berkeley (NJ)	<i>Ready, Set, Leap!</i>	18 classrooms	21 classrooms	T: 149 C: 137

¹ After one program or classroom attrited.

NOTE: T: Treatment Group

C: Control Group

Three research teams (Vanderbilt University, University of Texas Health Science Center at Houston, and Florida State University) have two treatment groups and a shared control group. When reading the "Students" column, the first "T" refers to the first curriculum in the same row, while the second "T" refers to the second curriculum in the same row. The "C" refers to the shared control group. For example, Vanderbilt University compared two curricula: *Bright Beginnings* (103 students) and *Creative Curriculum* (101 students) to a control curriculum (105 students).

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

dispersed to a large number of kindergarten classrooms in the second year of the study. Classroom observation data were collected in the fall and spring of the pre-kindergarten year using the Early Childhood Environment Rating Scale-Revised (ECERS-R) and the Arnett Caregiver Interaction Scale (Arnett) measures. The Teacher Behavior Rating Scale (TBRIS) was added to the classroom observation battery for the pre-kindergarten spring observation time point.

Outcome Variables

A total of 12 child-outcome measures and 11 intermediate outcome measures were used in the data analyses to answer the research questions outlined above. The measures included in the impact analyses are listed in table B-2. For the Woodcock Johnson (WJ) measures (Letter-Word Identification, Applied Problems, and Spelling subscales), the “W” scores were used in the analyses because the WJ standardized scores account for developmental growth associated with the child’s age in years. “W” scores are a special transformation of the Rasch ability scale. The “W” score for each test is centered on a value of 500, which has been set to the approximate performance of a 10-year-old student. Individuals whose performance on the measure is lower than the average score receive scores below 500. The use of the Rasch score allows researchers to record changes in actual ability within or across years of a study. For the Peabody Picture Vocabulary Test (PPVT), the Test of Early Reading Ability (TERA), the Test of Language Development (TOLD) Grammatical Understanding subtest, and the Pre-CTOPPP/CTOPP, Elision subtest measures, raw scores were used in the analyses. Standard scores were used for the SSRS measures, and standard T-scores were used for PLBS/LBS measures. T-scores are test scores converted to an equivalent standard score in a normal distribution with a mean of 50 and a standard deviation of 10.

Covariates

In all models used in the Main and Secondary analyses, covariates were included to increase the precision of the impact estimates by adjusting for chance baseline differences between the treatment and control groups on the characteristics represented by the covariates. In our models of child-outcome measures, the following set of covariates was used:

- child’s age;
- child’s race/ethnicity;
- child’s gender;
- self-reported maternal education;
- proxy disability status indicator (parent-reported IEP);
- site within research team (if a team had more than one site); and
- curriculum within research team (if a team evaluated more than one curriculum).

The following additional covariates were included in the repeated measures models:

- intervention exposure ($Time_1$) (i.e., time between start of intervention and the child’s assessment, in *all repeated measures models*), and
- post-intervention time ($Time_2$) (i.e., time since the spring assessment of the pre-kindergarten year in *repeated measures spline model only*).

The following additional covariate was included in the ANCOVA models:

- child’s baseline score on the relevant child-outcome measure.

Table B-2. Variables in analysis

Child outcome measures	Classroom outcome measures
<ul style="list-style-type: none"> • 3 mathematics measures: <ul style="list-style-type: none"> ○ WJ Applied Problems ○ CMA-A Mathematics Composite ○ Shape Composition¹ • 1 phonological awareness measure: <ul style="list-style-type: none"> ○ Pre-CTOPPP/CTOPP • 3 reading measures: <ul style="list-style-type: none"> ○ WJ Letter Word Identification ○ WJ Spelling ○ TERA • 2 language measures: <ul style="list-style-type: none"> ○ TOLD ○ PPVT • 3 behavior measures: <ul style="list-style-type: none"> ○ SSRS Social Skills ○ SSRS Problem Behaviors ○ PLBS/LBS 	<ul style="list-style-type: none"> • 3 classroom observation measures: <ul style="list-style-type: none"> ○ ECERS-R ○ Arnett Subscales <ul style="list-style-type: none"> ▪ Detachment ▪ Harshness ▪ Permissive ▪ Positive Interaction ○ TBRS: (Spring pre-kindergarten only) <ul style="list-style-type: none"> ▪ Book Reading ▪ Oral Language use ▪ Print and Letter Knowledge ▪ Written Expression ▪ Phonological Awareness ▪ Math Concepts

¹ Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

For the classroom models the following covariates were included:

- teacher has a BA degree;
- previous teaching experience;
- teacher’s race/ethnicity;
- child/adult ratio in classroom;
- average number of students attending classroom;
- city size;
- site within research team (if a team had more than one site); and
- curriculum within research team (if a team evaluated more than one curriculum).

The following additional covariate was included in the repeated measures models:

- intervention exposure (Time₁) (i.e., time between start of intervention and the classroom assessment, in all repeated measures models).

The following additional covariate was included in the ANCOVA models:

- baseline scores on the ECERS-R and Arnett measures were included in the models for these classroom outcomes.

Covariates Excluded From the Child-Level Analysis Models

Preliminary analyses were conducted to determine which covariates should be included in the child-outcome measures models. The analysis results indicated that the following list of covariates were unrelated to the child measures under consideration and accounted for little or no portion of the explained variance:

- child born in the United States;
- mother works full-time;
- two parents in the household;
- parents' income level;
- parents' age;
- primary language used in the home;
- mother reads daily;
- parents' health;
- mother depressed;
- child read to daily;
- number of types of children's reading material in home;
- child watches TV more than two hours per day;
- number of rules;
- number of weekly activities;
- number of monthly activities;
- teacher has AA degree;
- teacher has specialization in preschool;
- teacher has Child Development Associate (CDA) credential;
- classroom teaching experience;
- preschool teaching experience;
- teacher's salary level;
- teacher's age;
- teacher developmental attitudes;
- teacher didactic activities;
- child/adult ratio in classroom;
- number of child's absences;
- teacher reported extent of behavior problems in classroom;
- ECER-R total score; and
- Arnett total score.

Covariates Excluded from the Classroom-Level Analysis Models

Preliminary classroom level analyses looked at and eliminated the covariates listed below. As with the child covariates, these covariates were shown to be unrelated to classroom measures and accounted for little or no portion of the explained variance. The following covariates were not included in the final classroom models:

- whether or not the teacher has an AA degree;
- what specialization the teacher has;
- whether the teacher is credentialed in child development;
- classroom teaching experience;
- teacher’s salary level;
- developmental activities (composite from Teacher Interview);
- didactic activities (composite from Teacher Interview); and
- teacher reported extent of behavior problems of classroom.

Intervention/Control Group Assignment and Coding

An “intention-to-treat” logic¹ was employed for the site-level/curriculum-specific analyses of program impacts for child and classrooms measures. All children at a given site were included in the analysis in the group to which they were randomized, even if they were lost to follow-up for some reason.

To accurately compute differences between control and intervention groups, we utilized an intervention-within-sites effect-coding scheme. That is, for grantees with one site and one intervention, the intervention group was coded as +.05 and the control group was coded as -.05; for grantees with more than one site, dummy variables were created for each site that were coded +.05 and -.05 for intervention and control groups as above and coded 0 for information collected at the other site(s). This approach ensures that for sites that implemented only one curriculum, the significance of the intervention versus control difference is reflected in the parameter representing a particular intervention-within-site and that the control groups within sites were used as the comparison for their respective interventions. In cases where a grantee examined more than one intervention (i.e., Vanderbilt University, University of Texas Health Science Center at Houston [University of Texas-Houston] and Florida State University [FSU]) the effects of the shared control group were taken into account when reporting Intervention versus control differences. To obtain accurate estimates of the intervention effect, contrasts and estimates in SAS (Statistical Analysis Software) were used to provide the specified coefficients of interest.

Missingness—Attrition and Response Rates

Some attrition occurred in each of the 2 years in which data were collected, pre-kindergarten and kindergarten. The attrition rates were generally low and there was no evidence of differential attrition (i.e., attrition rates did not differ significantly between intervention and control groups). Response rates were lowest for the parent interview data collected by some research teams, and for the teacher reports at the end of the kindergarten year (see table B-3). The software used, SAS PROC MIXED, is designed to utilize all available data, even if some of the observations at a given time point are missing. In the two longitudinal models, if a child’s data were missing at one of the two follow-up data collection time points, the child’s earlier collected data were still included and used in estimation, where appropriate.

¹ “Intention to treat” is a strategy for the analysis of randomized controlled trials that compares individuals in the groups to which they were originally randomly assigned. It is based on the assumption that all of the individuals assigned to the intervention group may not receive the full exposure to the treatment, even though this was the original intent of the intervention.

Table B-3. Response rates and attrition

Research team	Response rate Fall 2003	Percent of sample with data Spring 2004	Percent of sample with data Spring 2005
Vanderbilt (n = 309)			
Child Assessments	100	94	97
Teacher Report	100	90	90
Parents Interview	82	81	75
UNC-Charlotte (n = 194)			
Child Assessments	98	88	85
Teacher Report	100	88	56
Parents Interview	87	69	71
University of New Hampshire (n = 123)			
Child Assessments	100	85	66
Teacher Report	99	81	50
Parents Interview	16	45	51
Success for All (n = 215)			
Child Assessments	98	95	90
Teacher Report	97	95	82
Parents Interview	91	94	86
University of Texas-Houston (n = 297)			
Child Assessments	99	94	79
Teacher Report	97	86	57
Parents Interview	80	74	68
University of North Florida (n = 244)			
Child Assessments	100	92	89
Teacher Report	96	89	64
Parents Interview	84	81	73
University of Virginia (n = 195)			
Child Assessments	85	96	97
Teacher Report	87	93	81
Parents Interview	93	87	89
Florida State University (n = 297)			
Child Assessments	95	96	80
Teacher Report	96	93	80
Parents Interview	91	84	75
UC-Berkeley and University at Buffalo, SUNY (n = 316)			
Child Assessments	99	94	90
Teacher Report	99	94	74
Parents Interview	83	90	78
Purdue and University of WI-Milwaukee (n = 204)			
Child Assessments	100	94	85
Teacher Report	100	90	66
Parents Interview	86	76	70

See notes at end of table.

Table B-3. Response rates and attrition—Continued

Research team	Response rate Fall 2003	Percent of sample with data Spring 2004	Percent of sample with data Spring 2005
University of Missouri-Columbia (n = 231)			
Child Assessments	99	90	81
Teacher Report	98	81	68
Parents Interview	92	84	84
UC-Berkeley (n = 286)			
Child Assessments	96	92	87
Teacher Report	96	95	84
Parents Interview	91	82	76
All teams (n = 2,911)			
Child Assessments	98	93	85
Teacher Report	97	90	72
Parents Interview	84	79	75

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Missingness—Item Non-response

To address item non-response in the parental interviews, we created missing status indicators for each of the parent interview categorical variables, to keep these cases in the analyses. This approach allowed us to fit a separate categorical effect for non-responders. The parameter estimates for the predictor variable associated with each missing data dummy variable is estimated based on available cases. This approach allowed for the inclusion of the maximum number of participants in all analyses and estimated associations between the predictors and outcomes based on all available cases.

Overall Modeling Framework

The general framework in which all of the analyses were conducted was the linear mixed effects model. This model generalizes the more familiar general linear model for regression by allowing for both random and fixed regression coefficients. The data analyses were conducted at the research team level. This was an appropriate approach because the selection and randomization of classrooms to treatment versus control condition were completed at the grantee site level for each intervention curriculum. Children were nested within classrooms in both the intervention and control groups and were repeatedly assessed with a battery of measures. These various sources of nesting (time within child and children within classrooms) were accounted for in the analyses. A hierarchical linear model (HLM) was therefore used for all study analyses to deal with the correlated (i.e., nested and repeated measures) data as well as the mixture of random and fixed effects in each model. The HLM model more accurately accounts for the variance in the data. This is important in order to obtain valid variance estimates and significance tests.

For child-outcome models, classroom was included as a random factor to account for the design effects associated with nesting of children within classrooms. We elected to test all treatment versus control inferences at a level generalizable to the entire population of similar students and classrooms/programs, which has the effect of increasing standard errors over inferences made only to the specific instances included

in the study. Models that explicitly modeled change over time (i.e., repeated measures and repeated measures spline models) included random effects for “time.”

The underlying assumptions for this model have been met, within reasonable limits. The main assumption is that each measure modeled is distributed according to a multivariate normal distribution. The linear mixed-effects model is relatively robust to some non-normality, such that extreme multivariate skewness or kurtosis would be needed to invalidate the model. It also assumes a linear relationship between the predictors and the outcomes. The model deals with the lack of independence among children in the same classroom by the inclusion of the random classroom intercept. The lack of independence due to repeated measures from the same children is addressed in the model by the inclusion of a covariance matrix for the errors that model this dependence among repeated measures.

Overview of Quantitative Modeling Utilized in the Study’s Primary and Secondary Analysis

As explained in the main report,² analysis on nearly all outcome measures was conducted in two different ways, one of which was considered primary and reported on in the main report, the other was considered secondary and was reported on only in appendix A. Within these two analysis approaches, models varied depending on the data structure of the longitudinal data (i.e., how many and at which time points data for a given outcome were collected). Tables B-4 and B-5 provide a summary overview of all the primary and secondary analyses presented in the main report and in appendix A.

Main Analysis—Reported in the Main Body of the Report

In the main report, we present repeated measures spline model analyses for data collected at three time points, simple repeated measures analyses for data collected at two time points, and ANCOVA analyses for data collected at one time point. We discuss these analyses below beginning with the repeated measures spline model, followed by a description of the simple repeated measures and ANCOVA analyses.

Secondary Analysis—Reported in Appendix A

In appendix A, we present additional analysis results from the repeated measures spline model analyses of data from three time points, analysis results from the simple repeated measures analyses of data from two time points, and ANCOVA analysis results of data from two or three time points. These secondary analyses were conducted to address questions related to nonequivalence at baseline and the possibility of early treatment effects.

² The “main report” refers to chapters 1-13.

Table B-4. Main analysis: Model used with each outcome measure

Outcome	Measure	Times observed	Model
Reading	TERA	3	Spline Repeated Measures
	WJ Letter Word Identification	3	Spline Repeated Measures
	WJ Spelling	3	Spline Repeated Measures
Phonological awareness ¹	Pre-CTOPPP	2	Simple Repeated Measures
	CTOPP	1	ANCOVA w/ Pre-K baseline
Language	PPVT	3	Spline Repeated Measures
	TOLD	3	Spline Repeated Measures
Mathematics	WJ Applied Problems	3	Spline Repeated Measures
	CMA-A Mathematics Composite	3	Spline Repeated Measures
	Shape Composition ²	3	Spline Repeated Measures
Pre-kindergarten behavior ¹	SSRS Social Skills	2	Simple Repeated Measures
	SSRS Problem Behaviors	2	Simple Repeated Measures
	PLBS	2	Simple Repeated Measures
Kindergarten behavior ¹	SSRS Social Skills	1	ANCOVA w/ Pre-K baseline
	SSRS Problem Behaviors	1	ANCOVA w/ Pre-K baseline
	LBS	1	ANCOVA w/ Pre-K baseline
Classroom quality	ECERS-R	2	Simple Repeated Measures
Teacher-child interaction	Arnett Detachment	2	Simple Repeated Measures
	Arnett Harshness	2	Simple Repeated Measures
	Arnett Permissiveness	2	Simple Repeated Measures
	Arnett Positive Interaction	2	Simple Repeated Measures
Literacy instruction	TBRS Written Expression	1	ANCOVA
	TBRS Print and Letter Knowledge	1	ANCOVA
Phonological instruction	TBRS Phonological Awareness	1	ANCOVA
Language instruction	TBRS Book Reading	1	ANCOVA
	TBRS Oral Language	1	ANCOVA
Mathematics instruction	TBRS Math Concepts	1	ANCOVA

¹ Pre-kindergarten and kindergarten measures not on the same scale.

² Building Blocks, Shape Composition task

NOTE: ANCOVA: Analysis of covariance. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten). Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B-5. Secondary analysis: Models and grades for each outcome measure

Outcome	Measure	Repeated Measures model	ANCOVA model with Pre-K baseline covariate
Reading	TERA	Spline: Pre-K and K	Pre-K and K
	WJ Letter Word Identification	Spline: Pre-K and K	Pre-K and K
	WJ Spelling	Spline: Pre-K and K	Pre-K and K
Phonological awareness ¹	Pre-CTOPPP CTOPP	Simple: Pre-K	Pre-K K
Language	PPVT	Spline: Pre-K and K	Pre-K and K
	TOLD	Spline: Pre-K and K	Pre-K and K
Mathematics	WJ Applied Problems	Spline: Pre-K and K	Pre-K and K
	CMA-A Mathematics Composite	Spline: Pre-K and K	Pre-K and K
	Shape Composition ²	Spline: Pre-K and K	Pre-K and K
Pre-kindergarten behavior ¹	SSRS Social Skills	Simple: Pre-K	Pre-K
	SSRS Problem Behaviors	Simple: Pre-K	Pre-K
	PLBS	Simple: Pre-K	Pre-K
Kindergarten behavior ¹	SSRS Social Skills		K
	SSRS Problem Behaviors		K
	LBS		K
Classroom quality	ECERS-R	Simple: Pre-K	Pre-K
Teacher-child interaction	Arnett Detachment	Simple: Pre-K	Pre-K
	Arnett Harshness	Simple: Pre-K	Pre-K
	Arnett Permissiveness	Simple: Pre-K	Pre-K
	Arnett Positive Interaction	Simple: Pre-K	Pre-K

¹ Pre-kindergarten and kindergarten measures not on the same scale.

² Building Blocks, Shape Composition task

NOTE: ANCOVA: Analysis of covariance. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten). Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Data Analysis Models

Repeated Measures

The repeated measures model provides a measure of intervention effects that is less influenced by differences in the variables under investigation that might exist between the two groups at baseline. Also, by including Time, the length of exposure to intervention, as a covariate, the model results (e.g., estimates of intervention group means or group differences at, for example, the pre-kindergarten spring time point) are less influenced by an early treatment effect that might exist at some sites. The inclusion of time also provides estimates of treatment and control group rates of change, allowing the differences in their rates of change to be tested.

Repeated Measures Linear Spline Model

For data collected in the fall and spring of pre-kindergarten and the spring of kindergarten, a repeated measures linear spline model was applied that modeled the three scores, using Time variables and a set of covariates. A simple graphical display of this model, also known as a “piece-wise linear” or “broken stick” model, is found in figure B-1. As the figure indicates, this model allows projection of the intervention and control group trajectories back to the start of the school year when curriculum implementation began. Group differences at the start of treatment can then be tested, assuming straight-line growth throughout the pre-kindergarten year.

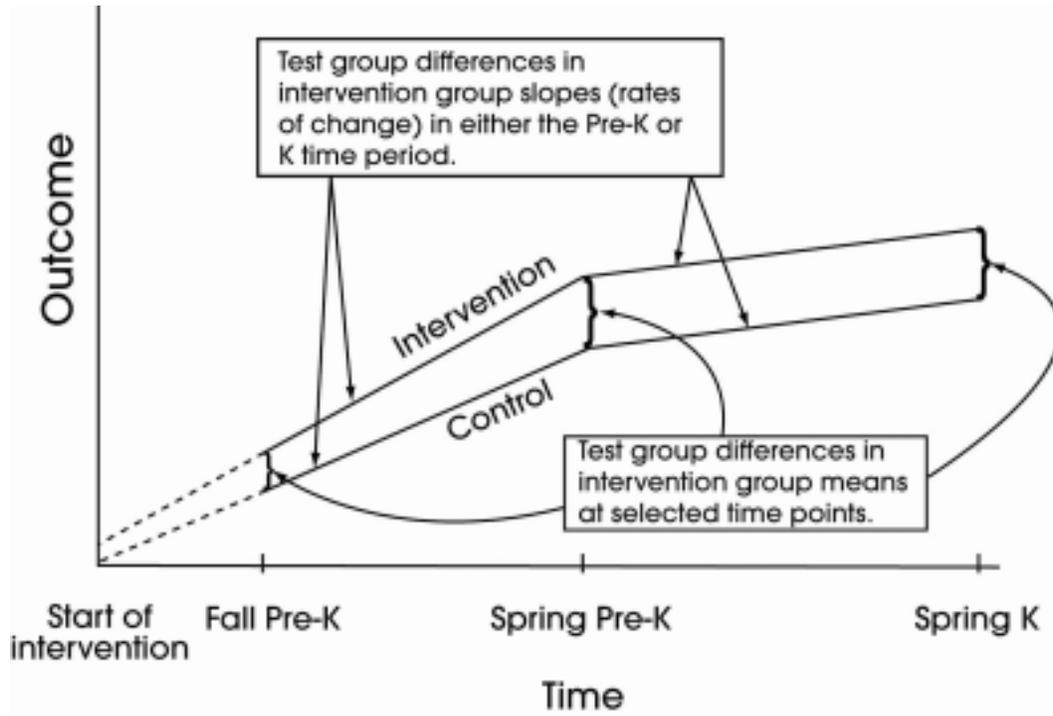
This model allows for the most extensive testing of the first two research questions: Question 1, regarding the child impacts at the end of preschool, and Question 2, regarding the child impacts at the end of kindergarten (some of which were sustained from preschool and others evident only at the end of kindergarten), and provides the most complete account of the data. Intervention-control group differences can be obtained at many points of interest: the spring pre-kindergarten and spring kindergarten time points, as well as baseline and start-of treatment time points. Group differences in the rate of change between intervention and control groups can be tested during the pre-kindergarten and kindergarten year (testing for “sleeper” and maintenance of intervention effects), as can within-group comparisons of the rates of change between the 2 years.

Simple Repeated Measures

For the four child outcomes and two classroom outcomes with comparable data from two time points, simple repeated measures analyses were conducted. This model accounts for repeated measures data by fitting a linear growth trajectory using the fall and spring time point means on the outcome of interest, adjusted for the covariates included in the model. Figure B-2 graphically displays this simple repeated measures model. As illustrated, the pre-kindergarten growth trajectories estimated by this model can be extended back to the start of treatment, allowing for a test of group differences at the start of treatment.

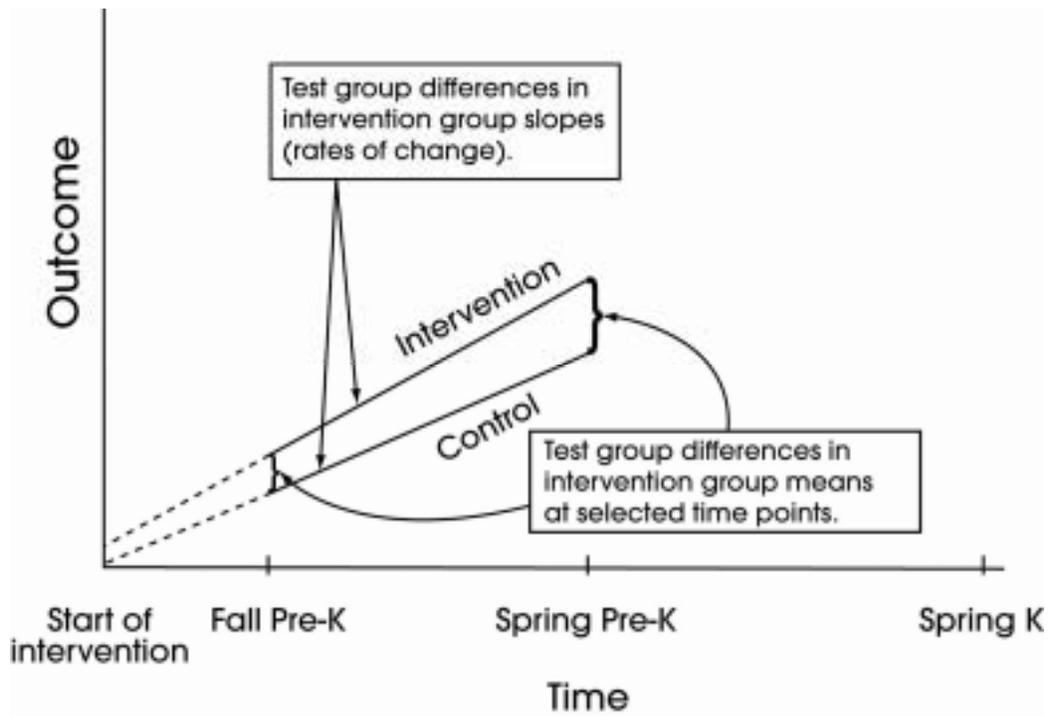
This model allows testing for intervention differences in spring scores adjusted for the covariates *and* for the amount of exposure to the intervention, as reflected in the Time variable. The model also affords parameter estimates to test the primary child research Question 1 (child outcomes at the end of pre-kindergarten) and research Question 3 (classroom outcomes at the end of pre-kindergarten). This model is similar to the repeated measures spline model described above, except that with two time points, it only models a single segment of the outcome trajectories covering the pre-kindergarten year.

Figure B-1. Repeated measures spline model



NOTE: Pre-K (pre-kindergarten); K (kindergarten)
SOURCE: The Preschool Curriculum Evaluation Research (PCER) study.

Figure B-2. Simple repeated measures model



NOTE: Pre-K (pre-kindergarten); K (kindergarten)
SOURCE: The Preschool Curriculum Evaluation Research (PCER) study.

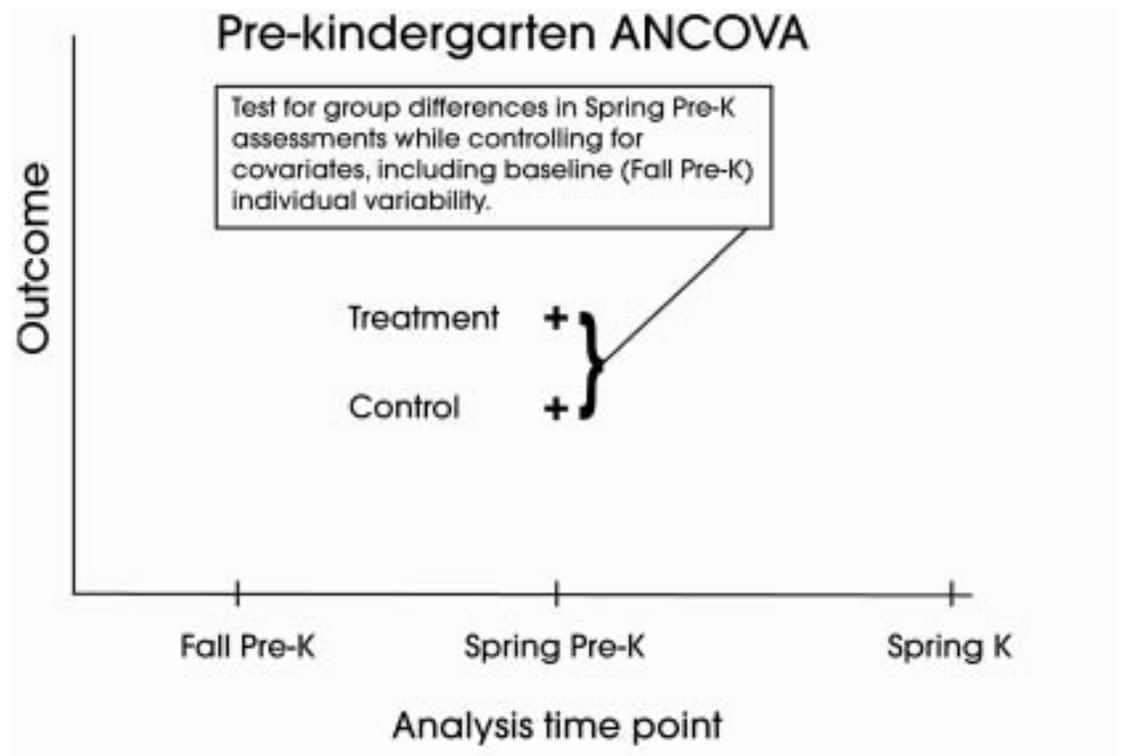
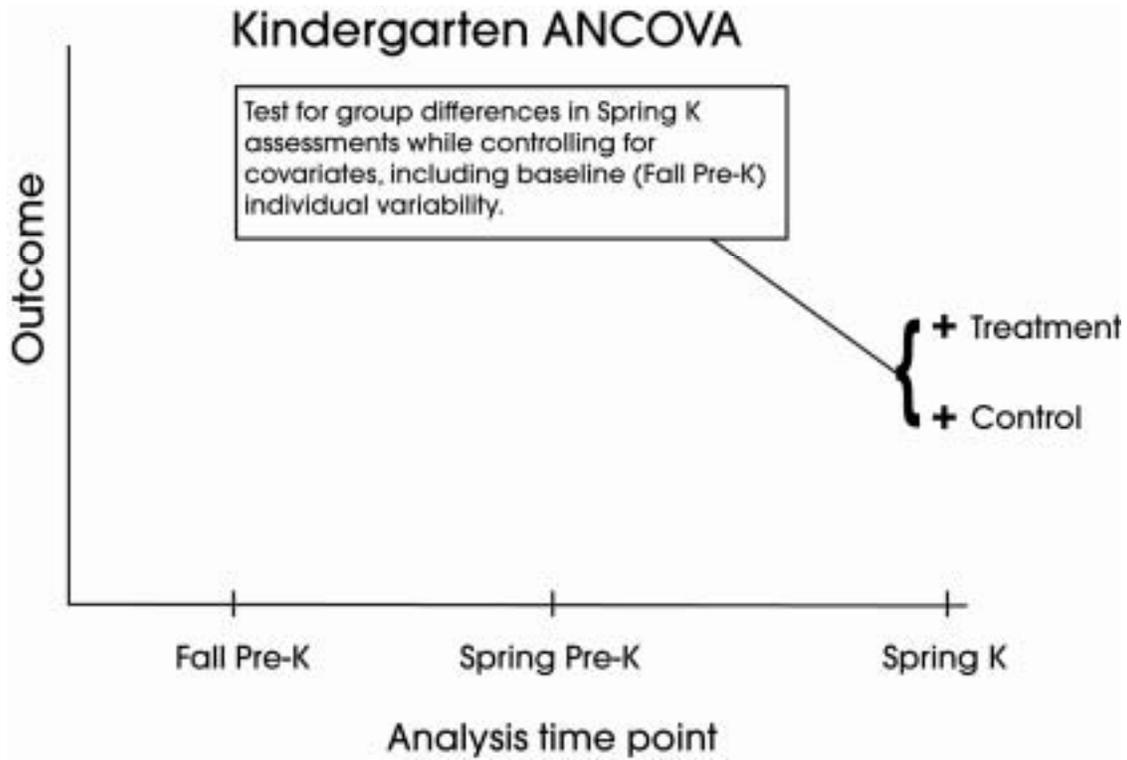
ANCOVA Model

For findings presented in the main report, an ANCOVA model was used to analyze comparable child or classroom data from one time point (see table B-4). In the ANCOVA analyses of child outcomes, the pre-kindergarten baseline score and the child and family covariates were included in the analysis model. For the ANCOVA analysis of classroom outcomes (TBRS classroom observation measure), the teacher, preschool, and community covariates were included in the model.

In appendix A, we present ANCOVA analyses for data from all time points. Two ANCOVA models, each modeling one or two spring outcomes using the child's fall pre-kindergarten score as a covariate, along with the child and family demographic covariates, were used to analyze child outcomes for which we have data from one, two, or three time points. We also present ANCOVA analyses for classroom outcomes with comparable data from one or two time points in appendix A.

Figure B-3 provides a graphical presentation of the ANCOVA model used with individual data for the child outcomes, and classroom data for the classroom outcomes. The ANCOVA model allows modeling of spring outcomes for both years, in spite of possible changes in measurement between either spring outcome and the baseline covariate. The ANCOVA model has a couple of disadvantages, namely, that (a) no rates of change can be estimated because the baseline is not modeled, and (b) this model may be biased by the possibility of an early treatment effect due to late baseline data collection, because the baseline assessment score is included in the model as a covariate. This model only provides estimates and tests of significance for group differences at the spring time point (end of pre-kindergarten or end of kindergarten) in each model. However, this is sufficient to provide answers to research Questions 1 and 2 for the child outcomes, and Question 3 for the classroom outcomes examined with this model.

Figure B-3. Pre-kindergarten (Pre-K) and kindergarten (K) analysis of covariance (ANCOVA) models



SOURCE: The Preschool Curriculum Evaluation Research (PCER) study.

Statistical Models

The analyses were conducted using the general mixed model framework. This model generalizes the general linear model (GLM) to include random as well as fixed effects, and allows for a more general variance-covariance structure for the covariates as well as error structure. Repeated measures and ANCOVA forms of the general mixed model were used to analyze the pre-kindergarten and kindergarten data.

General Multi-Level Model

A multi-level model (mixed-effect model) can be written in a matrix format as

$$\underline{Y} = X\underline{\beta} + Z\underline{u} + \underline{\varepsilon} \quad (1)$$

where

\underline{Y} is the n -dimension vector of observations, with n being the number of subjects in the study

$X_{n \times p}$ is the $n \times p$ design matrix for fixed effects, with p being the number of fixed effects

$\underline{\beta}_p$ is the p -dimension vector of fixed effect coefficients

$Z_{n \times r}$ is the $n \times r$ design matrix for the random effects, with r being the number of random effect parameters

\underline{u}_r is one vector of random effect parameters

$\underline{\varepsilon}_n$ is the n -dimension residual random error.

In this model, everything is the same as in the general linear model except for the addition of the known design matrix, Z , and the vector of unknown *random-effects parameters*, \underline{u} . The matrix Z can contain either continuous or dummy variables, just like X . The name *mixed model* comes from the fact that the model contains fixed effects parameters, $\underline{\beta}$, and random-effects parameters, \underline{u} . Henderson (1990) and Searle, Casella, and McCulloch (1992) provide a discussion of the historical developments of the mixed model.

If the covariance of the vector of random effects and the vector of error terms are given by

$$V(\underline{u}) = G \text{ and } V(\underline{\varepsilon}) = R \quad (2)$$

respectively, then the overall covariance structure of the observations is given as

$$V(\underline{Y}) = ZGZ' + R \quad (3)$$

A key assumption in the foregoing analysis is that \underline{u} and ε are normally distributed with

$$E \begin{bmatrix} \underline{u} \\ \underline{\varepsilon} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (4)$$

$$Var \begin{bmatrix} \underline{u} \\ \underline{\varepsilon} \end{bmatrix} = \begin{bmatrix} G & \Theta \\ \Theta & R \end{bmatrix}$$

One can obtain $V(\underline{Y})$ once both the random-effects design matrix Z and estimates from specified covariance structures for G and R are obtained. The covariance structures that can be used include variance components, unstructured, compound symmetry, and various time-dependent structures among others.

Repeated measures models

A repeated measures model, which is just a special case of the general model (1), was used to model assessments from the pre-kindergarten and kindergarten years simultaneously.

As an example, a model with only a random intercept term, if written in scalar format, is specified as

$$y_{ijk} = X_{ijk} \beta + u_i + \varepsilon_{ijk}, \quad i = 1, 2, \dots, r, \quad j = 1, 2, \dots, n_i, \quad k = 1, 2, 3$$

where r is the number of classrooms, n_i is the size of classroom i , (ij) indicates the j^{th} individual in the i^{th} classroom, u_i is the random intercept term, and k indicates the time point of the assessment (fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten). X_{ijk} may include a time effect, an intervention grouping variable, an interaction between intervention group and Time, along with other covariates. The Time variable was defined as the time since start of treatment for any given piece of assessment data. For the repeated measures linear spline model, X_{ijk} included two Time variables (Time₁ and Time₂). The first Time variable, Time₁, is the time since start of treatment. The second Time variable, Time₂, is the time since the pre-kindergarten spring assessment to the kindergarten assessment, and was defined as 0 (zero) for the pre-kindergarten assessments. The Time₂ variable provides a means of estimating any kindergarten-year increase or reduction in slope over and against the earlier pre-kindergarten-year slope that is represented by the coefficient for Time₁. The spline model included these two Time variables along with group interactions with each of the two Time variables: Intervention Group x Time₁ (for the pre-kindergarten period) and Intervention Group x Time₂ (for the kindergarten period). These interaction terms allowed us to include group-specific slopes for both time periods covered by the model.

When there was more than one random effect in the model (i.e., random intercepts and random slopes), a covariance structure among the random effects needed to be specified. Models treating Time₁ and/or Time₂ as random, in addition to the classroom intercept, were tested. In only one case was a random Time variable kept in the model (see table B-14, column regarding random variables). All grouping variables, covariates, and their interactions that were not treated as random were treated as fixed effects in the model.

Each child contributed three observations to the dataset used for model fitting. We imposed a correlation structure on observations from the same subject. Several different structures may be imposed, and we used a variety of structures to model this as appropriate (see Covariance Structure and Modeling Steps section for a description of the specific structures used in our analysis).

ANCOVA model

In the ANCOVA model for child outcomes, the fall pre-kindergarten (or baseline) measure was included in the model as a covariate, to increase precision by accounting for the variance that can be attributed to possible differences in children's scores on the baseline measures even with random assignment of groups. The child and family demographic covariates (i.e., child's age, race/ethnicity, gender, disability status, and maternal education) were also included in the model to increase precision of the estimates by accounting for any possible baseline differences on these characteristics. The model can be written in the form of equation (1), with one of the columns of X being the baseline score. Another column in X would be the site-specific or curricula-within-grantee-specific grouping variable, which captures the intervention impact. As an example, the following is a model used with a single random intercept and a correlated error term:

$$Y_{ij,t} = \beta_0 + u_i + \beta_1 Y_{ij,\text{baseline}} + X \beta + \varepsilon_{ij} \quad (5)$$

for the full model with covariates, where $Y_{ij,t}$ is the score at time t (say, spring kindergarten), $Y_{ij,baseline}$ is the fall pre-kindergarten score, $u_i \sim N(0, \sigma_{u^2})$, and $\varepsilon \sim N(0, R)$, for $j=1$ to n_i individuals, and $i=1$ to r classrooms.

The effects of interest in this analysis are the differences between the intervention and control group means. We used ESTIMATE statements in SAS PROC MIXED statistical software to provide adjusted means and two-tailed t-tests of intervention and control group mean differences at the start of intervention, fall, and spring assessments.

The confound between site and curriculum was handled by effect coding ‘‘Curriculum-in-Grantee’’ or ‘‘Site-in-Grantee.’’ Thus, we created separate variables for each curriculum at each site (e.g., *Pre-K Mathematics with DLM Early Childhood Express Math software package* in California, *Pre-K Mathematics with DLM Early Childhood Express Math software package* in New York, *Creative Curriculum* in North Carolina and Georgia) and coded control classrooms as -.5 and intervention classrooms as .5. Florida-FSU and the University of Texas-Houston research teams each had two intervention conditions that shared a control group; their intervention and control conditions were coded 1 and -1, respectively. Simple linear estimates and contrasts were used to specify model-predicted means at the fall and spring data collection points as well as model-predicted gain scores. This approach provided a parsimonious way of representing the design of the site-specific projects (e.g., a site with one intervention curriculum and one control group or a site with two intervention curricula and one control group).

Structures of Residual Covariance Matrix R

The covariance matrix of the residual errors for the repeated measures models are defined at the individual level, and takes the form of a block diagonal matrix. If we use R_{ij} to denote the covariance among observations from the j^{th} child in the i^{th} classroom, the $(i, j)^{th}$ diagonal block of R is defined for each model differently depending on the number of observations and the covariance structure we chose for the model. The notation (i, j) is not the typical $(i, j)^{th}$ element of a matrix, in this instance, all R_{ij} ’s are diagonal blocks of R . The residual covariance matrix is represented in the following form:

$$R = \begin{pmatrix} R_{n_1} & \dots & \Theta \\ \vdots & \ddots & \vdots \\ \Theta & \dots & R_{n_r} \end{pmatrix} \quad \text{and} \quad R_{n_i} = \begin{pmatrix} R_{i,1} & \dots & \Theta \\ \vdots & \ddots & \vdots \\ \Theta & \dots & R_{i,n_i} \end{pmatrix}$$

For the linear spline models with three time points, a heterogeneous compound symmetry (HCS) structure that assumes the covariance among residual errors of the same subject was used:

$$Cov(\varepsilon_{ij}) = R_{i,j} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 & \sigma_2\sigma_3\rho \\ \sigma_1\sigma_3\rho & \sigma_2\sigma_3\rho & \sigma_3^2 \end{pmatrix} \quad \text{for all } i, j$$

For the repeated measures models for data from two time points, an unstructured covariance structure was used and defined as

$$\text{Cov}(\xi_{ij}) = R_{i,j} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \quad \text{for all } i, j$$

For repeated measures models of the classroom outcomes (for data with two time points), a variance component covariance structure was used and defined as

$$\text{Cov}(\xi_{ij}) = R_{i,j} = \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_0^2 \end{pmatrix} \quad \text{for all } i, j$$

Structures of Random Effect Covariance Matrix G

For the linear spline models with three repeated measures, after experimenting with different structures for the random effects (including both the random intercept and slopes), a model with only a random intercept term was settled on:

$$G = \sigma_u^2 I, \text{ and } Z = \begin{bmatrix} \mathbf{1}_{n_1} & & & \\ & \mathbf{1}_{n_2} & & \\ & & \ddots & \\ & & & \mathbf{1}_{n_r} \end{bmatrix}$$

where $\mathbf{1}_{n_i}$ is a vector of 1's of dimension n_i .

When modeling child outcomes with comparable data from two observations (i.e., behavioral outcome models, see Covariance Structure and Modeling Steps for more details), a random slope seemed justified. Therefore, those models include a random slope term, as well a random intercept term. The structure of G and Z are more complicated and defined as

$$G = \begin{pmatrix} \Sigma & \dots & \Theta \\ \vdots & \ddots & \vdots \\ \Theta & \dots & \Sigma \end{pmatrix} \text{ and } Z = \begin{bmatrix} Z_{n_1} & & & \\ & Z_{n_2} & & \\ & & \ddots & \\ & & & Z_{n_r} \end{bmatrix}$$

where G and Z are a block diagonal matrices of the same dimensions. The diagonal blocks of G are identical and are given as

$$\Sigma = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_s^2 \end{pmatrix}$$

with σ_u^2 , and σ_s^2 being the variance components for the random intercept and random slope, respectively, and Z_{n_i} being matrices of size $2 \times n_i$ with the first column containing all 1's and second column containing the observed values of Time (time interval between the start of intervention and the particular follow-up assessment) for each subject in classroom i :

$$\begin{bmatrix} 1 & T_{i,1} \\ \vdots & \vdots \\ 1 & T_{i,n_i} \end{bmatrix}$$

Assumed Variance Structure of Outcomes

After making assumptions regarding covariance matrices G and R , the assumed overall covariance structure of the outcome under modeling can be derived from the following relationship:

$$V(Y) = ZGZ' + R$$

As an example, the assumed covariance structure for the repeated measures models with both a random intercept, a random slope, and observations from two time points is given as

$$Cov(y_{ijk}, y_{i'j'k'}) = \begin{cases} \sigma_u^2 + T_{i,j}^2 \sigma_s^2 + \sigma_1^2 & \text{if } i = i', j = j', k = k' = 1 \\ \sigma_u^2 + T_{i,j}^2 \sigma_s^2 + \sigma_2^2 & \text{if } i = i', j = j', k = k' = 2 \\ \sigma_u^2 + T_{i,j}^2 \sigma_s^2 + \sigma_{12} & \text{if } i = i', j = j', k \neq k' \\ \sigma_u^2 + T_{i,j} T_{i,j'} \sigma_s^2 & \text{if } i = i', j \neq j' \\ 0 & \text{if } i \neq i' \end{cases}$$

where i is the index for classroom, j for individual child, and k for time point. The structure implies that observations from children of different classrooms are independent. Observations from children of the same classroom are correlated and the covariance is accounted for by the intra-classroom covariance σ_u^2 and the covariance of the random classroom slope σ_s^2 . Observations from the same child are correlated, and the covariance depends on both the intra-classroom correlation and the correlation among observations of the same child (as specified in the residual covariance matrix R).

Mean Model and Testing of Fixed Effects

To illustrate this mixed model approach we use a slightly simplified linear spline model (excluding covariates) with two groups (intervention and control). Two fully developed examples from our analyses are found in the section Group Comparisons Testing for Intervention Impact. The model with the Time₁, Time₂, Group, and their attendant interaction terms is as follows:

$$E(Y_{ij}) = \beta_1 + \beta_2 \text{Time}_1 + \beta_3 \text{Time}_2 + \beta_4 \text{Group}_1 + \beta_5 (\text{Time}_1 \times \text{Group}_1) + \beta_6 (\text{Time}_2 \times \text{Group}_1)$$

For this model, when the control group is coded as 0 (zero), its expectation is

$$E(Y_{ij}) = \beta_1 + \beta_2 \text{Time}_1 + \beta_3 \text{Time}_2$$

whereas the intervention group which is coded as one, has as its expectation:

$$E(Y_{ij}) = (\beta_1 + \beta_4) + (\beta_2 + \beta_5) \text{Time}_1 + (\beta_3 + \beta_6) \text{Time}_2$$

Hypotheses tested included:

$$H_0 : \beta_4 = 0 \text{ (No group difference at start of intervention),}$$

$$H_0 : \beta_5 = 0 \text{ (No group difference in pre-kindergarten growth rate),}$$

$$H_0 : \beta_5 + \beta_6 = 0 \text{ (No group difference in kindergarten growth rate),}$$

$$H_0 : \beta_3 = \beta_6 = 0 \text{ (No change in the slope in the kindergarten year).}$$

In the analyses, effect coding is used rather than dummy coding; therefore, the betas have a somewhat different meaning than that illustrated above. With effect coding, the control group is coded -1 and the treatment group is coded 1 (or some variant of this coding) to account for the various grantee design configurations. This coding scheme results in the regression parameters being equal to treatment effects obtained in traditional analysis of variance models. For example, with effect coding, β_1 represents the overall mean, and β_4 represents the effect of the treatment relative to the control group (Kirk 1995). As with the example provided above, the contrasts to test for intervention/control group differences in mean levels and slopes were all estimable and conducted with the effect coding used in the model specification.

Sample Weights and Missing Data

Sample weighting to account for variability across sites in sample sizes and intervention-control balance is unnecessary in the mixed model. The mixed model approach used for the ANCOVA or repeated measures spline model accounts for the unbalanced design when analyses are conducted using SAS PROC MIXED. The PROC MIXED analysis is valid with unbalanced data and missing data, as long as they are missing randomly (Littell et al. 1996). No contrary evidence was found to this assumption.

Estimation Methods

As indicated above, the mixed model that was used must estimate parameters in four matrices: (1) the variance-covariance matrix for the random effects, G ; (2) the error structure for the residual error, R ; (3) the fixed effects parameters, β ; and (4) the random-effects parameters, y . The elements of G and R are estimated using restricted, or residual, maximum likelihood (REML). Fixed parameters are estimated using the Maximum Likelihood (ML) method. SAS PROC MIXED was used to carry out the estimation. PROC MIXED allows for three types of estimation: ML, REML, or MIVQUE0 (Minimum Variance Quadratic Unbiased Estimation). Swallow and Monahan (1984) showed that REML and ML variance component estimates are superior to MIVQUE0. As reported in Swallow and Monahan (1984), under some circumstances, MIVQUE0 is a poor estimator of unbalanced data. They also note that the ML estimates have the smallest MSE (means square error); however, this is due to the downward bias in these estimates. The REML-estimated MSEs are larger precisely because the estimates are unbiased. REML estimates are unbiased and considered superior to ML estimates because they account for the degrees of freedom lost in the estimation of the fixed effects. We further discuss the relative merits of REML and ML estimation in the next section. The Proc Mixed estimation method was implemented using a sweep-based Newton-Raphson algorithm (Wolfinger et al. 1994).

For the fixed model estimates, you have to obtain estimates of β and y the standard method is to solve the mixed model equations (Henderson 1984):

$$\begin{bmatrix} X\hat{R}^{-1}X & X\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} X\hat{R}^{-1}y \\ Z'\hat{R}^{-1}y \end{bmatrix} \quad (6)$$

The solutions can also be written as

$$\begin{aligned}\hat{\beta} &= (X\hat{V}^{-1}X) - X\hat{V}^{-1}y \\ \hat{\gamma} &= \hat{GZ}\hat{V}^{-1}(y - X\hat{\beta})\end{aligned}\tag{7}$$

and have connections with empirical Bayes estimators (Laird and Ware 1982). These estimates are obtained in SAS PROC MIXED using maximum likelihood (ML) estimation.

When estimates for G and R are used to estimate β and y , the resulting estimates have favorable statistical properties. The estimate of β is considered to be the empirical best linear unbiased estimator (EBLUE) of β , and the estimate of y is considered an empirical best linear unbiased predictor (EBLUP) of y .

REML Versus ML Estimation Method

SAS PROC MIXED offers two methods of maximum likelihood estimation of the variance components. One is the full maximum likelihood (FML or ML), and the other is the restricted (or residual) maximum likelihood (REML or RML). Both use the likelihood principle but differ with regard to how the likelihood function is constructed.

The full ML estimates of the variance components contain ML estimates of the fixed effects when estimating variance components. This method treats the estimated fixed effects as known. By ignoring the uncertainty due to the fact that the fixed effects are sample estimates, ML estimates of the variance components overstate the degrees of freedom available for estimation and therefore underestimate the variance of the fixed effect parameters. Such concerns led to the development of REML (Patterson and Thompson 1971; Harville 1974; Dempster, Laird, and Rubin 1977). REML maximizes a residual version of the likelihood function, after the estimation of the fixed effects. As a result, the variance component estimates obtained from the two approaches can be numerically different. The point can be illustrated using an example of an ordinary regression model (with no random effects), assuming we have a total of N independent observations and if we fit an ordinary regression model with p parameters (fixed effects). If we further denote the true variance of the observations as σ^2 , then the ML and the REML versions of the estimate of variance have the following relationship with the true variance:

$$E(\hat{\sigma}_{ML}^2) = \frac{(N-p)}{N} \sigma^2, \quad \text{and} \quad E(\hat{\sigma}_{REML}^2) = \sigma^2\tag{8}$$

Thus, the REML estimate of the variance is unbiased and adjusts the ML estimate of variance by the following multiplier:

$$\frac{N}{N-p} = \frac{\text{Number of observations}}{\text{Number of observations} - \text{Number of fixed effect parameters}} \geq 1\tag{9}$$

For random effect models, both ML and REML estimates are calculated through iterative procedures and an equivalent adjustment is made. The REML likelihood function is obtained by multiplying the usual ML likelihood function by a factor that is the square root of the generalized variance of the fixed effect parameters.

To understand the impact of this adjustment, we again use the above example of ordinary regression model. The multiplier defined in (9) is close to 1 when the number of fixed parameters is much smaller than the number of observations. However, when the difference of the two is small, it could be quite different from one. The number of fixed effect parameters used in our models differs by model; however, they are the same

for a given model across all sites and outcomes. For example, the linear spline model described in the section on the repeated measures models has 21 fixed effect parameters and the ANCOVA model described in the section ANCOVA model has 16 fixed effect parameters. The number of observations used in fitting those models to child outcomes ranges between 300 and 100, with the majority in the lower portion of the range nearer 100. The smaller sites, however, had sample sizes as low as 50 (NH ANCOVA model). For the classroom outcomes models, the number of observations included in the models is even smaller. As a result, the adjustment factor (9) can sometimes be substantially larger than 1. That is, ML estimates can underestimate the true value by a substantial amount.

Although REML is supposed to be less biased than ML when sample sizes are small, it is not guaranteed to produce numerically superior estimates in all circumstances (Kreft and deLeeuw 1998). One reason for this is that REML estimates may become more variable than ML estimates as the variance of the fixed effect parameters becomes larger (or the sample size becomes smaller), thus increasing the possibility of REML behaving worse than ML. Again, using the ordinary regression model in (8) as an example, if $N = 30$ and $p = 20$, then REML estimates would have degrees of freedom of 10 versus 30 for the ML estimates. As a result, although the expected REML estimates are still unbiased, the probability of the REML estimate of a particular sample being further away from the true value of the parameter is much higher than the corresponding ML estimate.

As one cannot know for sure whether this less desirable property of REML is being manifested in any particular situation, the odds of that occurring must be balanced against the almost certain bias that occurs with ML estimates. The consensus is that REML is theoretically the approach of choice. Dempster, Laird, and Rubin (1977) declared REML to be “intuitively more correct.” In addition, REML solutions have the desired property of being equivalent to the ANOVA estimators for balanced models. Most standard texts on mixed models (Verbeke and Molenberghs 2000; McCulloch and Searle 2001; Fitzmaurice, Laird, and Ware 2004) advise using REML, and consider it the preferred approach. As a result, we used REML estimates for all reported significance levels.

Early Intervention Effects and Estimated True Baseline

The start of baseline data collection began 2 weeks or more after the start of the school year (start of intervention curriculum implementation) at several research sites. This data collection timeline raised concerns among members of the PCER Consortium that early effects of the experimental curriculum may have already occurred by the time the baseline assessments were collected. Failure to obtain true baselines (by which we mean measures of child and classroom data uncontaminated by the start of intervention) may result in biased estimated growth rates and growth rate impacts during the pre-kindergarten year, with the direction of that bias undetermined. It is difficult to ascertain whether this situation occurred because we did not collect data from children prior to the start of the school year.

To address concerns related to early treatment effects due to the absence of a “true” baseline, we included a time variable that reflects the amount of exposure to the intervention in the repeated measures models (see *section on Repeated Measures Models*). This time variable takes a value of 0 (zero) at the initiation of the intervention and measures all other assessment time points from this origin. Although the estimated baseline group means may still have been affected by early intervention effects, with this covariate in the model, the estimated rate of change from repeated measure models is much less influenced by potential early treatment effect due to delays in the baseline assessment.

Using this time variable, group means were estimated at time 0 (zero) (i.e., the start of the intervention) for both the intervention and control groups even when the baseline assessment was conducted weeks later. These means were estimated the same way the group means were estimated for other time points in the model. The means were estimated by just setting Time_1 and Time_2 equal to 0 (zero) in the spline model or just $\text{Time} = 0$ in the simple repeated measures model. Significance testing of the extrapolated group differences at

time 0 (zero) allows us to conclude, under the model assumptions, whether the groups differed at the start of intervention (beginning of the school), not just at baseline (time of the fall pre-kindergarten assessment). This set of estimated means is not problem-free, as it relies upon an assumption that the growth process is relatively linear from start of intervention through the spring assessment. To the extent this assumption does not hold, particularly in the period of time before the baseline assessment, the start-of-intervention estimates should be interpreted with caution. Despite the possible limitations, extrapolating back to the start of the intervention and estimating treatment and control groups mean differences for the start of treatment and fall baseline assessment was our attempt to address issues related to the possibility of early treatment effects.

The results across all 14 curricula indicated that there were very few instances in which there were statistically significant differences between the intervention and control groups at start of intervention. This suggests, if the linear growth assumption holds, that there is very little evidence of imbalance between the intervention and control groups in terms of the child outcomes at the start of the study, after adjusting for the standard set of covariates. However, in some cases (e.g., classroom-level models for grantees with small numbers of classrooms) these tests have low power to detect statistically significant differences.

Multiple Comparisons

No adjustments were made to the p -values from the results of this study before comparing them to the significance levels, indicated in the report by asterisks. Multiple outcomes were examined with the same children, teachers, and classrooms, and within each outcome multiple time points were examined and multiple contrasts tested. Therefore, the overall “family-wise” type I error rate is not protected at a particular significance level by only considering results significant at that level or lower. If a conservative line aimed at protecting against type I errors at all costs were adopted, the power to detect effects would decline. Arguments have been made by some that such adjustments are not necessary for a variety of reasons (Feise 2002; Rothman 1990; Savitz and Olshan 1995). Among these reasons is the suggestion that in some studies power considerations are more important than type I errors. The nature of this evaluation was more exploratory, and sought significant effects for curricula measured against an already high bar found in most prevailing evaluations. The reader may not wish to put too much weight on the p -values, although significance levels are reported for the results. The possibility of spurious results are possible, but some protection against them was afforded by the decision not to accept a curriculum as having made a significant difference in a domain unless it was found to do so on more than one measure in that domain. Also, the focus on effect sizes provides a better sense of the amount of improvement afforded by the various intervention curricula on each measure.

Effect Size Calculation

Definition of Effect Size

The significant intervention effects are determined by tests of linear combinations of beta coefficients estimated for the model; the reflection of that difference is reported as an effect size in the main report and in appendix A. Effect sizes are often used to provide a relative measure of the magnitude of differences. In this report, the effect size is defined as the difference between two statistics of interest divided by a normalizing factor:

$$d = \frac{\delta\mu}{\sigma} = \frac{\mu_2 - \mu_1}{\sigma} \quad (10)$$

The statistics of interest, μ_i 's, could be means of experimental groups at a particular time point (such as the mean of an intervention group at fall pre-kindergarten), differences in group means at two different time

points, as well as slopes of the growth curves. Both the numerator (the difference) and denominator (the normalizing factor) need to be estimated from the sample data.

The function of the normalizing factor is to remove the influence of population variation from the measure of interest. Many quantities can be used for this factor: the estimated standard deviation (SD) of the control group at baseline or an average of the estimated standard deviations of the two groups at baseline. When the groups are created from the same population using proper randomization procedures, the underlying population variation for them is the same and pooling estimates from the two groups increases the reliability of the estimated SD.

Calculation of the Numerator

The numerators for the effect size calculations are differences of means of control and intervention groups at various time points and differences in slopes of the growth curves of different groups between various time points. Nevertheless, they can all be expressed in terms of linear combinations of parameters of the models we fit, either repeated measures or ANCOVA models. As such, the general formula for estimating an effect size is given as

$$\hat{d} = \frac{C' \beta}{\hat{\sigma}_{pooled}} = \frac{\beta_2 - \beta_1}{\hat{\sigma}_{pooled}} \quad (11)$$

where β 's are the model parameters and σ_{pooled} the pooled estimated population standard deviation as described above.

For comparisons of means, either child or classroom outcomes, the difference in the two betas shown in (11) becomes the difference between two means. For example, for the spring pre-kindergarten comparison of control vs. intervention, using $\hat{M}_{Pre-K,S,C}$ as the estimated mean for the control group at the spring pre-kindergarten assessment and $\hat{M}_{Pre-K,S,T}$ the estimated mean for the intervention group, and σ_{pooled} the appropriate pooled estimated standard deviation for the outcome of interest (either child or classroom), is defined as:

$$\hat{d}_{Pre-K,S,means} = \frac{\hat{M}_{Pre-K,S,T} - \hat{M}_{Pre-K,S,C}}{\hat{\sigma}_{pooled}} \quad (12)$$

Similarly, for child or classroom outcomes in slope differences, Cohen's d for the pre-kindergarten slope comparison of control vs. intervention at the Pre-k, spring time point can be defined as follows. Let $\beta_{pre-k,S,C}$ be the pre-kindergarten slope for the control population at the spring data collection, $\beta_{Pre-K,S,T}$ be the same for the intervention population, and again σ_{pooled} is the pooled estimated standard deviation for whichever child or classroom outcome is of interest. The effect size for the difference in slopes for the pre-kindergarten year would be:

$$\hat{d}_{Pre-K,S,Slopes} = \frac{\beta_{Pre-K,S,T} - \beta_{Pre-K,S,C}}{\hat{\sigma}_{pooled}} \quad (13)$$

Choice of the Normalizing Factor

For the PCER study data, the statistics of interest for which effect sizes are calculated include differences between the means of control and intervention groups at various time points, and differences between the slopes of the growth curves of control and intervention groups between various time points. As such, the different statistics of interest are estimated, using data from different time points. In determining the most appropriate normalizing factor for the calculation of the effect sizes, besides requiring that it reflect the variation of the population under study, a common approach for all outcomes and, for each outcome, a common normalizing factor that can be used for all appropriate statistics and for as many sites as possible (except sites with different designs and outcomes with fewer time points) was sought. This reduced the complexity of the calculation and made it easier to understand, but also provides a unified platform on which results are compared. In addition, using a uniform approach made significant results less likely due to happenstance.

One possible argument against this approach is that since the knowledge and skills of the young children change over time, so does the reference population. Thus, it is necessary to investigate change in the population variation. For this purpose, the population variation was estimated by time point for each of the outcomes. After averaging the estimated population standard deviations across group assignment and site, the results are summarized in table B-6.

Table B-6. Average estimated population standard deviation by study outcome and time points

Outcome	Fall Pre-K	Spring Pre-K	Spring K	Average
PPVT	16.56	16.76	15.09	16.13
Pre-CTOPPP	3.64	4.12	—	3.88
CMA-A Mathematics Composite	0.23	0.23	0.17	0.21
TOLD	4.77	4.64	3.86	4.42
Shape Composition ¹	0.85	0.87	0.80	0.84
WJ Letter Word Identification	25.09	23.96	28.61	25.89
WJ Applied Problems	21.51	18.42	17.24	19.06
WJ Spelling	25.93	25.39	22.43	24.58
TERA	6.88	8.98	9.81	8.56
SSRS Social Skills	14.87	14.17	—	14.52
SSRS Problem Behaviors	13.02	12.82	—	12.92
PLBS	10.08	10.26	—	10.17
Average	11.94	11.72	12.25	11.94

— Not available.

¹ Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

As can be seen, for most outcomes, only small changes occur in population variation between the first two time points (fall pre-kindergarten and spring pre-kindergarten). There are indications, however, that the population variation becomes smaller at the third time point (spring kindergarten). One possible explanation is that children perform more uniformly once they are in a structured school setting such as kindergarten. The results also suggest that population variation of the first two times should be considered as the benchmark for population variation. To increase the reliability of the sample estimates, the weighted average of the estimated population variation across group assignment and the first two time points as the normalizing factor in the calculation of the effect sizes was used.

To assess the impact of pooling the estimated population variation between the control and intervention groups, the variance estimates between the two groups were investigated. Averaging across site, and all available time points, the results are summarized in table B-7.

Again, for most of the outcomes, when averaging over site and time point, the difference between estimated population standard deviation of the control and intervention groups was minimal.

Table B-7. Average estimated population standard deviation by study outcome and group assignment

Outcome	Control group	Intervention group	Average
PPVT	16.78	15.59	16.13
Pre-CTOPPP	3.92	3.85	3.88
CMA-A Mathematics Composite	0.22	0.21	0.21
TOLD	4.57	4.30	4.42
Shape Composition ¹	0.86	0.82	0.84
WJ Letter Word Identification	26.30	25.53	25.89
WJ Applied Problems	20.31	17.98	19.06
WJ Spelling	24.70	24.48	24.58
TERA	8.72	8.41	8.56
SSRS Social Skills	14.32	14.69	14.52
SSRS Problem Behaviors	12.96	12.89	12.92
PLBS	10.36	10.01	10.17
Average	12.20	11.71	11.94

¹ Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures. When there are two intervention groups, the number for intervention is the average of the two standard deviations.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Calculation of the Normalizing Factor

The analyses focused on a single outcome measure at a time. Modeling and estimation were carried out site by site for each intervention curriculum. Thus, for each measure and site combination, a normalizing factor needed to be obtained, requiring two steps: (1) determine the appropriate data points to be included for the calculation, and (2) the actual calculation of the pooled standard deviation. Table B-8 provides an example of how the data were selected for this calculation under a typical situation: a grantee with a single site and a single intervention where data from three time points are available. Cells used in estimating the pooled population standard variation are indicated as “used in pooling”. Others are not used in the calculation.

Table B-8. Pooled standard deviation example

Intervention or control	Fall Pre-K	Spring Pre-K	Spring K
Intervention	Used in pooling	Used in pooling	Not used in pooling
Control	Used in pooling	Used in pooling	Not used in pooling

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

This could be data on a Woodcock-Johnson measure from a single site grantee such as New Hampshire. The variances and sample sizes for the data of the four cells (intervention and control groups for fall pre-kindergarten and spring pre-kindergarten assessment data) are pooled, using a weighted combination (see below) to arrive at the pooled standard deviation. The pooled standard deviation obtained with these data is used to calculate effect size estimates for comparisons made between control and intervention groups for the longitudinal spline model as well as the two ANCOVA models of spring pre-kindergarten and kindergarten outcomes. There are many exceptions to the typical example illustrated in table B-8. For example, there are sites where more than one intervention is used. There are also outcome measures where data from only two time points are available. The section entitled Pooled Standard Deviation Calculation Details for Each Data Structure by Research Team Site/Intervention Configuration, provides further details on the data selection step for grantees with more complicated designs and outcomes for which we collected data at different time points.

Once the data points used for pooling are determined, calculation of the pooled standard deviation is carried out by fitting a null model (with no covariates) with a classroom-level random effect. For example, for each cell in table B-8 above, a model as follows is fit:

$$y_{i,j} = \mu_0 + u_i + \varepsilon_{i,j} \quad (14)$$

where i, j are indices for classroom and children within a classroom, respectively; $y_{i,j}$ is the outcome under study; μ_0 the intercept term; and u_i and $\varepsilon_{i,j}$ are the classroom random effect and individual residual error, respectively. The resulting estimated variance components from fitting model (14) are the classroom variance component, $\hat{\sigma}_u^2$ associated with u_i 's and the residual variance, $\hat{\sigma}_\varepsilon^2$, associated with $\varepsilon_{i,j}$'s. The sum of the two components provides the total population variation for the outcome/site/intervention/time point combination.

$$\hat{\sigma}_{total, trt, k}^2 = \hat{\sigma}_{u, trt, k}^2 + \hat{\sigma}_{\varepsilon, trt, k}^2 \quad (15)$$

where trt and tp are indicators for intervention group assignment and time point, respectively.

Next, the pooled population variance estimate is calculated as a weighted average of estimates (15) from all cells as

$$\hat{\sigma}_{Pooled}^2 = \left(\sum_{trt,k}^K (n_{trt,tp} - 1) \hat{\sigma}_{Total, trt, k}^2 \right) / \left(\sum_{trt,k}^K (n_{trt,k} - 1) \right) \quad (16)$$

where K is the total number of cells created by the intervention by time point combination for the particular outcome and site. Because the sample sizes within each cell are different, the weighting by relative sample sizes produces an unbiased estimate of the population variance.

The method of estimating population variance used here can be validated by examining the sample variance without considering the hierarchical nature of the sample. Table B-9 shows the population standard deviation for twelve child outcomes as estimated using SAS PROC MEANS compared with estimates using the mixed model (14), as described above. The results are remarkably close. Since estimates from SAS PROC MEANS do not take into account the hierarchical structure of the sample data, those estimates were consistently slightly smaller than estimates from the mixed model, validating the additional variance components that have been accounted for by the results from the mixed model (14).

Table B-9. Estimated pooled population standard deviation using unconditional standard deviations and standard deviations from repeated measures analyses

Outcome	Mixed model	SAS PROC MEANS
PPVT	17.3826	17.2059
Pre-CTOPPP	3.9495	3.9210
CMA-A Mathematics Composite	0.2377	0.2355
TOLD	4.8631	4.8343
Shape Composition ¹	0.8948	0.8887
WJ Letter Word Identification	25.4034	25.0996
WJ Applied Problems	21.5521	21.3774
WJ Spelling	25.9873	25.6834
TERA	8.2939	8.2063
SSRS Social Skills	15.0638	14.8298
SSRS Problem Behaviors	13.3030	13.1097
PLBS	10.5641	10.4620

¹ Building Blocks, Shape Composition task

NOTE: The standard deviations were calculated using SAS PROC MEANS and Mixed Model averaged over sites. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Pooled Standard Deviation Calculation Details for Each Data Structure by Research Team Site/Intervention Configuration

This section provides additional details on data pooling for the calculation of the normalization factor used in reporting the effect sizes. This study features a variety of designs (four in all) among the PCER study grantees. The tables (table B-10 through table B-12) generalize table B-8 to cover the various situations encountered in our analysis, both in terms of design and the time points at which data on an outcome were collected. For each of our team site/intervention configurations, the tables contain a row for each intervention group within the given configuration. Within these rows, the cells (group by time period combinations) of data used in estimating a given pooled population standard variation for a particular site or intervention within a team are indicated by the inclusion of the same individual letter (A, B, or C) in that set of cells. “A” for site 1 or intervention 1, “B” for site 2 or intervention 2, et cetera. Teams with two or three sites had results reported at the individual site level and combined across the grantees sites. Pooling for such combined results for double or triple site grantees are indicated by letter combinations (e.g., AB or ABC). Those cells not used in any calculation for that set of outcomes and analyses are marked with an “x.” Cells for which no such data exist are marked with a “-.” Thus, for example, for a child outcome with three time points of data at a single site/double intervention (such as Vanderbilt University, University of Texas-Houston, and Florida State University), the “A”s in four of table B-10’s cells of the appropriate intervention 1 and control rows under the “Intervention 1 Test” columns indicate that these four cells were used in the calculating the pooled standard deviation used for estimating the effect size for intervention 1. Similarly, the four cells that contain “B”s in the rows for intervention 2 and control are pooled to form the pooled standard deviation for estimating effect sizes for intervention 2. Another example is the case of a research team working at two sites with a single intervention (i.e., the University of California, Berkeley with the University at Buffalo, State University of New York (California/New York) working in both California and New York) with data at all three time points. The “A”s could indicate pooling for the California site’s pooled standard deviation used in calculating their effect sizes, the “B”s pooling for the New York site’s pooled standard deviation and effect size calculations, and the “AB”s indicate cells involved in pooling for combined California/New York pooled standard deviations used in effect size estimations. Table B-10 contains the pooling used for the bulk of the analysis, the spline models, repeated measures with pre-kindergarten data, and pre-kindergarten and spring

kindergarten ANCOVAs for which pre-kindergarten measures were the same as kindergarten measures. Table B-11 contains the pooling for the ANCOVAs with TBRS outcomes. Table B-12 contains the pooling for spring kindergarten ANCOVAs on measures that changed between pre-kindergarten and kindergarten assessments.

Table B-10. Pooled standard deviation details: Pooling for outcomes modeled with the simple repeated measures, the repeated measures spline models, the pre-kindergarten spring analysis of covariance (ANCOVA) models (except Teacher Behavior Rating Scale (TBRS)), and the kindergarten spring ANCOVA models where kindergarten data were comparable to pre-kindergarten

Intervention and control group combinations	Fall Pre-K		Spring Pre-K		Spring K
Single site/single intervention					
Intervention	A		A		X
Control	A		A		X

Intervention and control group combinations	Fall Pre-K		Spring Pre-K		Spring K
	Intervention 1 test	Intervention 2 test	Intervention 1 test	Intervention 2 test	
Single site/double intervention					
Intervention 1 Sites	A	†	A	†	X
Intervention 2 Sites	†	B	†	B	X
Control	A	B	A	B	X

Intervention and control group combinations	Fall Pre-K		Spring Pre-K		Spring K
	Single site tests	Combined site tests	Single site tests	Combined site tests	
Double site/single intervention					
Intervention group-Site 1	A	AB	A	AB	X
Control group-Site 1	A	AB	A	AB	X
Intervention group-Site 2	B	AB	B	AB	X
Control group-Site 2	B	AB	B	AB	X

Intervention and control group combinations	Fall Pre-K		Spring Pre-K		Spring K
	Single site tests	Combined site tests	Single site tests	Combined site tests	
Triple site/single intervention					
Intervention group-Site 1	A	ABC	A	ABC	X
Control group-Site 1	A	ABC	A	ABC	X
Intervention group-Site 2	B	ABC	B	ABC	X
Control group-Site 2	B	ABC	B	ABC	X
Intervention group-Site 3	C	ABC	C	ABC	X
Control group-Site 3	C	ABC	C	ABC	X

† Not applicable.

NOTE: A: Intervention group-Site 1, Control group-Site 1
 B: Intervention group-Site 2, Control group-Site 2
 C: Intervention group-Site 3, Control group-Site 3
 X: Data were not used in the calculations.

Details about reading this table are found in the section Pooled Standard Deviation Calculation Details for Each Data Structure by Team Site/Intervention Configuration. A simpler example is illustrated in table B-8. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten).

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B-11. Pooled standard deviation details: Pooling for Teacher Behavior Rating Scale (TBRS) outcomes modeled with the pre-kindergarten spring analysis of covariance (ANCOVA) models

Intervention and control				
group combination	Fall Pre-K	Spring Pre-K		Spring K
Single site/single intervention				
Intervention	—	A		—
Control	—	A		—
Intervention and control				
group combination	Fall Pre-K	Spring Pre-K		Spring K
		Intervention 1 test	Intervention 2 test	
Single site/double intervention				
Intervention 1	—	A	†	—
Intervention 2	—	†	B	—
Control	—	A	B	—
Intervention and control				
group combination	Fall Pre-K	Spring Pre-K		Spring K
		Single site tests	Combined site tests	
Double site/single intervention				
Intervention group-Site 1	—	A	AB	—
Control group-Site 1	—	A	AB	—
Intervention group-Site 2	—	A	AB	—
Control group-Site 2	—	A	AB	—
Intervention and control				
group combination	Fall Pre-K	Spring Pre-K		Spring K
		Single site tests	Combined site tests	
Triple site/single intervention				
Intervention group-Site 1	—	A	ABC	—
Control group-Site 1	—	A	ABC	—
Intervention group-Site 2	—	B	ABC	—
Control group-Site 2	—	B	ABC	—
Intervention group-Site 3	—	C	ABC	—
Control group-Site 3	—	C	ABC	—

— Not available.

† Not applicable.

NOTE: A: Intervention group-Site 1, Control group-Site 1

B: Intervention group-Site 2, Control group-Site 2

C: Intervention group-Site 3, Control group-Site 3.

Details about reading this table are found in the section Pooled Standard Deviation Calculation Details for Each Data Structure by Team Site/intervention Configuration. A simpler example is illustrated in table B-8.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B-12. Pooled standard deviation details: Pooling for kindergarten spring outcomes (SSRS, Pre-CTOPPP/CTOPP, PLBS/LBS) modeled with analysis of covariance (ANCOVA) models

Intervention and control group combination	Fall Pre-K	Spring Pre-K	Spring K	
Single site/single intervention				
Intervention	—	—	A	
Control	—	—	A	
Intervention and control group combination	Fall Pre-K	Spring Pre-K	Spring K	
Single site/double intervention				
Intervention 1	—	—	A	†
Intervention 2	—	—	†	B
Control	—	—	A	B
Intervention and control group combination	Fall Pre-K	Spring Pre-K	Spring K	
Double site/single intervention				
Intervention group-Site 1	—	—	A	AB
Control group-Site 1	—	—	A	AB
Intervention group-Site 2	—	—	B	AB
Control group-Site 2	—	—	B	AB
Intervention and control group combination	Fall Pre-K	Spring Pre-K	Spring K	
Triple site/single intervention				
Intervention group-Site 1	—	—	A	ABC
Control group-Site 1	—	—	A	ABC
Intervention group-Site 2	—	—	B	ABC
Control group-Site 2	—	—	B	ABC
Intervention group-Site 3	—	—	C	ABC
Control group-Site 3	—	—	C	ABC

— Not available.

† Not applicable.

NOTE: A: Intervention group-Site 1, Control group-Site 1

B: Intervention group-Site 2, Control group-Site 2

C: Intervention group-Site 3, Control group-Site 3

Details about reading this table are found in the section Pooled Standard Deviation Calculation Details for Each Data Structure by Team Site/intervention Configuration. A simpler example is illustrated in table B-8. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Model Specification

Standardization of Covariates

To enhance interpretability of model estimates, all continuous covariates were standardized (mean = 0, standard deviation = 1). Two covariates were standardized in the child outcome models: the child's age and the fall baseline measure. Thus, child outcome model estimates are for a child at the mean value of all continuous covariates and an average of the effects for all categories for each categorical covariates (e.g., average age, average fall baseline, averages across effects for child's gender, race/ethnicity, maternal education and parent reported individual education plan categories). For the classroom outcome models, the four continuous covariates that were standardized include previous years of preschool teaching, child-to-adult ratio, average class size, and the fall baseline measure. The standardization of these covariates means the classroom outcome estimates are specified for a classroom considered to have an average child to adult ratio, class size, and fall baseline score, and taught by a teacher with an average amount of preschool teaching experience and equally averaged across the categories of the classification covariates: city size, the teacher's BA attainment, and teacher's race.

Data Clustering—Nesting of Children in Kindergarten Classrooms and Blocking

Analysis of educational data such as these in which children reside in fixed classrooms must take into account the common experience they share that tends to reduce their individual variance. The data clustering introduced by the use of intact classrooms can be included by introducing such nesting into the model. As indicated above, our model incorporated this by estimating random classroom intercepts. In these multi-site longitudinal hierarchical data there were two possible nesting structures (represented by the random classroom intercepts) that could have been used in the analyses: the pre-kindergarten and kindergarten classroom nestings. The nesting of children in the pre-kindergarten classrooms was by study design (i.e., multiple children typically were selected from pre-kindergarten study classrooms), whereas nesting in the kindergarten classrooms reflected the extent to which children in the pre-kindergarten programs attended the same or different primary schools. In addition, intervention was implemented in the pre-kindergarten classrooms. Therefore, it was logical to use the pre-kindergarten classrooms as the nesting structure in our models.

We examined the clustering of the data in the pre-kindergarten and kindergarten years separately by intervention group and site. Table B-13 displays the range of cluster sizes for our data in the pre-kindergarten- and kindergarten-year classrooms. Though not provided in the table, a general description of the distributions of these sites by intervention-group cluster sizes is provided below. The distributions in the kindergarten year are all highly right skewed with the majority of the classrooms having a frequency of one child per class, and much lower frequencies for any larger class clusters. The data in the pre-kindergarten year are more variable, though practically never right skewed, and then only slightly. The data distribution is generally flat or unimodal, with a few being left skewed (lower frequencies for smaller class sizes and higher frequencies for larger class sizes). In most cases the pre-kindergarten minimum cluster size is larger than the kindergarten maximum cluster size, indicating the relative rareness of kindergarten year clustering. In virtually every case of overlap in the cluster size distributions (i.e., the maximum kindergarten class size is larger than the minimum pre-kindergarten class size), the overlap results from a single classroom, which tends to be an individual outlier among the rest of the classes within that site by intervention group in that year.³ Fully 37

³ For example, the site 4 control group shows some overlap, but only one pre-kindergarten classroom had only 5 children and the next smallest classroom had 10 children. Only one classroom in kindergarten had six and the next largest had only four. Site 15 has the most overlap. In the pre-kindergarten year the control group had one classroom with one student and one classroom with two. All the others had four or more. In the kindergarten year, the site had one classroom each with five, four, and three children, and all the others are two or less. Similarly in the intervention group, in the pre-kindergarten year only one classroom each had two and four children clustered in it, while in the kindergarten year only one each had eight, six, or four children clustered in the same classrooms.

percent of the kindergarten clusters were singletons (having only one PCER child in the classroom) and another 20 percent were clusters of two. In pre-kindergarten 95 percent of the clusters were larger than five while in kindergarten only 9 percent were this large. It seemed clear that the pre-kindergarten clustering was preferred for this analysis, as cluster sizes in kindergarten are so small, especially relative to the pre-kindergarten clusters, and have only a minimal impact on the estimation of error variability. In fact, within each team/site, very often multiple preschools are in the study. However, since we consider the classroom to be the most influential clustering structure, it is the only structure reflected in the model.

Table B-13. Pre-kindergarten and kindergarten classroom clusters of children, maximum and minimum size

Site	Group	Pre-kindergarten		Kindergarten	
		Maximum	Minimum	Maximum	Minimum
1	Control	9	7	5	1
1	Intervention	8	8	4	1
2	Control	9	6	6	1
2	Intervention	8	7	4	1
3	Control	10	4	4	1
3	Intervention	14	7	6	1
4	Control	16	5	6	1
4	Intervention	14	3	4	1
5	Control	7	5	2	1
5	Intervention	10	8	4	1
6	Control	11	6	7	1
6	Intervention	16	4	4	1
7	Control	14	11	3	1
7	Intervention	13	12	4	1
8	Control	11	8	3	1
8	Intervention	11	8	2	1
9	Control	19	8	7	1
9	Intervention	19	10	8	1
10	Control	7	3	5	1
10	Intervention	9	5	5	1
11	Control	17	10	12	1
11	Intervention	17	16	11	1
12	Control	21	5	7	1
12	Intervention	20	2	5	1
13	Control	14	7	8	1
13	Intervention	15	5	6	1
14	Control	8	5	4	1
14	Intervention	8	5	4	1
15	Control	8	1	5	1
15	Intervention	10	2	8	1
16	Control	11	6	2	1
16	Intervention	15	3	3	1
17	Control	13	12	10	1
17	Intervention	13	8	4	1
18	Control	17	9	11	1
18	Intervention	17	12	7	1

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

We also should note that there is also site-level blocking that occurs in the site data, as randomization takes place within sites. The team-level analysis that we conducted takes this into account with site-specific effect coding whenever multiple sites were involved for a given team.

Covariance Structure and Modeling Steps

To account for the clustering mentioned above as well as the correlated data for individual children or classrooms across time points, a mixed model was used which features means of accounting for these sources of covariance. The covariance structure for each of the models was determined in the initial steps of conducting those analyses, one for random effects in the model accounting for the classroom clustering and the other for the correlations among the repeated measures.

In determining the covariance structure of the random effects, there were three random effects were considered in the linear spline model (intercept, pre-kindergarten slope, and kindergarten slope), two in the simple repeated measures model (intercept and pre-kindergarten slope), and a single random intercept in the ANCOVA models of child outcomes. Preliminary model fitting looked at both a variety of structures, beginning with a completely unstructured covariance that included random effect variances for the intercept, pre-kindergarten slope, and kindergarten slope and their correlations. Whether there was sufficient randomness found to justify including each of these effects in the model as random was checked. These latter checks were done using nested likelihood tests with critical points from the appropriate 50:50 mixture of chi-squared distributions (Fitzmaurice, Laird, and Ware 2004). The former were examined by considering comparative values for appropriate information criteria (Akaike's Information Criterion [AIC], Bayesian Information Criterion [BIC], and the corrected Akaike's Information Criterion [AAIC]). In the vast majority of cases, the variance estimate for the classroom intercept random effect was significant, indicating that if not accounted for, the total variation would be underestimated. Random pre-kindergarten and kindergarten slopes were generally not statistically significant and, except for the repeated measures model for the child behavioral outcomes, were not included as random effects in the final model. The random classroom intercept was included in all repeated measures and ANCOVA models for individual child outcomes. For classroom outcome models classroom clustering is not an issue since classrooms are the level of analysis, thus no random effects were included in those models.

The most appropriate covariance structures for the residual variances across the time points were identified using nested model comparisons when possible and information criteria when nesting was not possible. This was done by starting with unstructured models then examining a wide variety of structures that appeared possible, based on the estimates from the unstructured models. In general, we found that several covariance structures fit the data equally well (i.e., the data did not strongly indicate that one structure was much better than any other). Resulting covariance structures used for each final model are detailed in table B-14 and their definitions are given in table B-15.

The covariance structure was chosen and the mean model structure refined, where necessary, using nested model comparisons with maximum likelihood (ML) estimation. Finally, variance components were estimated using REML for the final model estimation, providing unbiased results for testing of group differences and other results tabled in this report. A more detailed discussion concerning REML versus ML estimation is given in the section on Estimation. These analyses were conducted at the research team level, with separate estimates for sites and curricula within sites for each team.

Table B-14. Specific covariance structures found to best fit the data

Variables analyzed	Analysis model	Random variables and any covariance structure (G)	Repeated Measures residual covariance structure (R)
Analyses reported in the main report¹			
Eight child cognitive outcomes/ three time points	Repeated Measures Spline Model	Classroom intercept	Heterogeneous compound symmetry
Cognitive outcome/two time points	Simple Repeated Measures	Classroom intercept	Unstructured
Behavioral outcomes/two time points	Simple Repeated Measures	Classroom intercept and time/variance components	Unstructured
Classroom outcomes/one time point	ANCOVA	None	Simple residual variance
Analyses reported in appendix A			
Eight child cognitive outcomes/ three time points	Repeated Measures Spline Model	Classroom intercept	Heterogeneous compound symmetry
Cognitive outcome/two time points	Simple Repeated Measures	Classroom intercept	Unstructured
Behavioral outcomes/two time points	Simple Repeated Measures	Classroom intercept and time/variance components	Unstructured
Eight child cognitive outcomes/ three time points	ANCOVA	Classroom intercept	Simple residual variance
Cognitive outcome/two time points	ANCOVA	Classroom intercept	Simple residual variance
Behavioral outcomes/two time points	ANCOVA	Classroom intercept	Simple residual variance
Classroom outcomes/two time points	ANCOVA	None	Simple residual variance

¹ The term main report refers to chapters 1-13.

NOTE: ANCOVA: Analysis of covariance. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten).

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B-15. Specific covariance structure definitions

Covariance structure	Definition
Heterogeneous compound symmetry	The (i,j) th element of the covariance matrix is defined as: $\sigma_i \sigma_j 1(i \neq j) + 1(i = j)$
Variance components	The (i,j) th element of the covariance matrix is defined as: $\sigma_k^2 1(i = j)$ and i corresponds to the k th effect. This means the variances for the random intercepts and random slopes were each freely estimated, and the covariance between them is fixed at zero.
Unstructured	The (i,j) th element of the covariance matrix is defined as: σ_{ij} This structure allows the model to freely estimate the random variance for the two time points, as well as the covariance between them.

¹ The formalization, $1(i \neq j)$, assigns 1 to this term if $i \neq j$; the formalization $1(i = j)$, assigns 1 to the term if $i = j$.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Test for Equality of Pre-kindergarten and Kindergarten Period Slopes

The linear spline model estimates separate intervention group growth trajectories for the pre-kindergarten instructional period (projected back to the start of the intervention and tested for differences in the tables) and for the period following the pre-kindergarten instructional period, ending in the spring of the kindergarten year. If the rate of change during the two time periods is the same, then this model does not need to include a separate slope term for the kindergarten period. To check for such a possible model simplification, equality of slopes for the preschool and kindergarten years was tested. This simplification did not fit the data well; indicating rates of growth did in fact differ across the pre-kindergarten and kindergarten periods.

Site by Covariate Interactions

Since all modeling was done at the research team and curriculum level, site by covariate interactions are ruled out in most cases. These were checked for teams with multiple sites. The vast majority of these were nonsignificant, and thus these terms were not included in the models.

Homogeneity of Regression Assumption

As part of our preliminary model checking, the equality of the covariates' slopes were checked for equality across intervention groups by including intervention by covariate interaction terms. Table B-16 contains the significant interactions that were found. While this number of significant interactions may seem large, this is only 10 percent of the 840 possible interactions. Only one factor, disability status, emerged somewhat consistently, accounting for 23 percent of the "significant" interactions. However, the very small number of children with disabilities in many intervention or control groups resulted in rather odd and unstable interactions. Therefore, all intervention x covariate interactions in subsequent analyses were excluded.

Group Comparisons⁴ Testing for Intervention Impact

All intervention impact comparisons were conducted at the curriculum level, because each research team selected and implemented its intervention independently. Accordingly, intervention differences in means and rates of change over time had to be tested at the team level. This means all data for a given team's sites were analyzed in the same model, with terms included to separate out effects for different curricula or sites for teams with two intervention curricula, or two geographic locations and one intervention curriculum. As stated previously, the repeated measures models allowed a simultaneous test of whether intervention differences in the spring score(s), and the rates of change between specified time points (e.g., fall-spring pre-kindergarten or spring pre-kindergarten to kindergarten). These means and rates of change were estimated across sites for grantees with multiple geographic locations (e.g., the combined effect of *Pre-K Mathematics with DLM Early Childhood Express Math software* package in California and New York). Statistical tests for mean or slope intervention impacts were conducted using t-tests with estimated degrees of freedom, as explained below. All tests were two-tailed tests and the significance level was indicated with asterisks at .05, .01, .001, and .0001 significance levels (see tables with the impact analysis results in the main report and appendix A).

⁴ Group comparisons refer to treatment versus control group comparisons for each intervention.

Table B-16. Significant treatment by covariate interactions from check on homogeneity of regression assumption

Variable	Effect	Prob F
TERA	ELLM_BAY * Child'sAge	0.0306
TERA	ELLM_BAY * DisabilityProxyFall	0.0028
TERA	ELLM_JCK * Child'sGender	0.0258
TERA	CC(UNC)_NC * DisabilityProxyFall	0.0296
TERA	BB_TN * Child'sAge	0.0164
TERA	PA_WI * Child'sAge	0.0430
TERA	PC_MO * DisabilityProxyFall	0.0360
TERA	LE_FL * Child'sAge	0.0249
TERA	DLM with OC_FL * Child'sRace	0.0161
TERA	Curiosity Corner_FL * MaternalEducation	0.0258
TERA	Curiosity Corner_NJ * DisabilityProxyFall	0.0029
WJ Applied Problems	ELLM_JCK * Child'sGender	0.0182
WJ Applied Problems	ELLM_MIA * Child'sAge	0.0340
WJ Applied Problems	CC (UNC)_NC * MaternalEducation	0.0318
WJ Applied Problems	CC(UNC)_NC * DisabilityProxyFall	0.0019
WJ Applied Problems	LB_TX * DisabilityProxyFall	0.0137
WJ Applied Problems	PC_MO * DisabilityProxyFall	0.0008
WJ Applied Problems	LE_FL * DisabilityProxyFall	0.0417
WJ Applied Problems	Curiosity Corner_KS * Child'sAge	0.0497
WJ Applied Problems	Curiosity Corner_KS * DisabilityProxyFall	0.0198
WJ Applied Problems	Curiosity Corner_NJ * MaternalEducation	0.0147
WJ Applied Problems	Curiosity Corner_NJ * DisabilityProxyFall	0.0022
Shape Composition ¹	ELLM_JCK * Child'sAge	0.0454
Shape Composition ¹	CC(UNC)_NC * DisabilityProxyFall	0.0106
Shape Composition ¹	LE_FL * DisabilityProxyFall	0.0033
Shape Composition ¹	Curiosity Corner_FL * Child'sGender	0.0108
Shape Composition ¹	Curiosity Corner_FL * Child'sAge	0.0018
Shape Composition ¹	Curiosity Corner_NJ * MaternalEducation	0.0030
Shape Composition ¹	LFC_VA * Child'sGender	0.0204
WJ Letter Word Identification	Pre-K Math_CA * Child'sAge	0.0087
WJ Letter Word Identification	Pre-K Math_NY * MaternalEducation	0.0331
WJ Letter Word Identification	ELLM_BAY * DisabilityProxyFall	0.0369
WJ Letter Word Identification	ELLM_MIA * DisabilityProxyFall	0.0036
WJ Letter Word Identification	CC(UNC)_NC * DisabilityProxyFall	< 0.0001
WJ Letter Word Identification	BB_TN * Child'sAge	0.0303
WJ Letter Word Identification	DD_TX * DisabilityProxyFall	0.0470
WJ Letter Word Identification	PC_MO * DisabilityProxyFall	0.0079
WJ Letter Word Identification	Curiosity Corner_KS * DisabilityProxyFall	0.0008
WJ Letter Word Identification	Curiosity Corner_FL * MaternalEducation	0.0124
WJ Letter Word Identification	Curiosity Corner_NJ * MaternalEducation	0.0205
WJ Letter Word Identification	Curiosity Corner_NJ * DisabilityProxyFall	0.0156

See notes at end of table.

Table B-16. Significant treatment by covariate interactions from check on homogeneity of regression assumption—Continued

Variable	Effect	Prob F
WJ Spelling	Pre-K Math_CA * DisabilityProxyFall	0.0274
WJ Spelling	ELLM_BAY * DisabilityProxyFall	0.0044
WJ Spelling	CC(UNC)_NC * DisabilityProxyFall	0.0102
WJ Spelling	CC(UNC)_TN * Child'sRace	0.0310
WJ Spelling	CC(UNC)TN * DisabilityProxyFall	0.0147
WJ Spelling	BB_TN * Child'sRace	0.0261
WJ Spelling	LB_TX * DisabilityProxyFall	0.0008
WJ Spelling	DD_TX * DisabilityProxyFall	0.0047
WJ Spelling	PC_MO * Child'sGender	0.0476
WJ Spelling	PC_MO * DisabilityProxyFall	0.0066
WJ Spelling	LE_FL * Child'sAge	0.0006
WJ Spelling	LE_FL * DisabilityProxyFall	0.0416
WJ Spelling	Curiosity Corner_KS * DisabilityProxyFall	< 0.0001
WJ Spelling	Curiosity Corner_FL * Child'sGender	0.0082
WJ Spelling	Curiosity Corner_FL * MaternalEducation	0.0332
WJ Spelling	Curiosity Corner_NJ * MaternalEducation	0.0497
PPVT	ELLM_MIA * Child'sAge	0.0060
PPVT	ELLM_MIA * Child'sRace	0.0411
PPVT	CC(UNC)_NC * DisabilityProxyFall	0.0011
PPVT	CC(UNC)_TN * Child'sGender	0.0376
PPVT	PA_WI * Child'sRace	0.0122
PPVT	Curiosity Corner_KS * Child'sGender	0.0162
PPVT	Curiosity Corner_FL * MaternalEducation	0.0291
TOLD	Pre-K Math_CA * Child'sAge	0.0305
TOLD	ELLM_MIA * Child'sAge	0.0054
TOLD	CC(UNC)_NC * DisabilityProxyFall	0.0414
TOLD	LB_TX * DisabilityProxyFall	0.0177
TOLD	DD_TX * DisabilityProxyFall	0.0348
TOLD	PA_WI * Child'sGender	0.0151
TOLD	LE_FL * DisabilityProxyFall	0.0029
TOLD	Curiosity Corner_FL * MaternalEducation	0.0012
TOLD	Curiosity Corner_NJ * MaternalEducation	0.0055

See notes at end of table.

Table B-16. Significant treatment by covariate interactions from check on homogeneity of regression assumption—Continued

Variable	Effect	Prob F
CMA-A Mathematics Composite	Pre-K Math_CA * Child'sGender	0.0158
CMA-A Mathematics Composite	Pre-K Math_CA * DisabilityProxyFall	0.0102
CMA-A Mathematics Composite	Pre-K Math_NY * Child'sGender	0.0114
CMA-A Mathematics Composite	CC(UNC)_NC * Child'sAge	0.0461
CMA-A Mathematics Composite	CC (UNC)_NC * DisabilityProxyFall	0.0153
CMA-A Mathematics Composite	LB_TX * DisabilityProxyFall	0.0079
CMA-A Mathematics Composite	PA_WI * Child'sGender	0.0237
CMA-A Mathematics Composite	LE_FL * Child'sAge	0.0005
CMA-A Mathematics Composite	LE_FL * DisabilityProxyFall	0.0052
CMA-A Mathematics Composite	DLM with OC_FL * Child'sAge	0.0241
CMA-A Mathematics Composite	Curiosity Corner_KS * DisabilityProxyFall	0.0052
CMA-A Mathematics Composite	Curiosity Corner_NJ * MaternalEducation	0.0414

¹ Building Blocks, Shape Composition task

NOTE: Refer to the glossary for abbreviations of the measures. Abbreviations for the curricula are:

BB: *Bright Beginnings*

CC(UNC): *Creative Curriculum* (University of North Carolina at Charlotte)

DD: *Doors to Discovery*

DLM with OC: *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

ELLM: *Early Literacy and Learning Model*

LB: *Let's Begin with the Letter People*

LE : *Literacy Express*

LFC : *Language-Focused Curriculum*

PA: *Project Approach*

PC: *Project Construct*

Pre-K Math: *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Example for a single intervention at a single site

An example of testing intervention differences using the linear spline model for child outcomes is described here for the University of New Hampshire research team that implemented one curriculum (*Creative Curriculum* with *Ladders to Literacy*) at one site (NH). In this example, y is the child outcome under modeling; Time_1 is the time lapses between the start of treatment and any subsequent assessment used in the model; Time_2 is the time lapse between the pre-kindergarten spring assessment and later assessments, respectively, as described in the section Repeated Measures Models; and $\text{CTRL} - \text{NH}$ is the intervention by site indicator.

$$\begin{aligned}
 E(y) = & \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3 (\text{CTRL-NH}) \\
 & + \beta_4 (\text{CTRL-NH} \times \text{Time}_1) + \beta_5 (\text{CTRL-NH} \times \text{Time}_2) \\
 & + \beta_6 \text{Gender} + \beta_7 \text{Age} + \beta_8 \text{Race} + \beta_9 \text{Disability} \\
 & + \beta_{10} \text{MaternalEducation}
 \end{aligned}$$

The evaluation includes estimating means for the intervention and control classrooms at a particular time, typically spring of pre-kindergarten year. The intervention by site indicators are coded in such a way that linear contrasts can be conveniently constructed to allow comparisons of groups means and slopes at various time points. With this coding scheme, the group means, adjusted for covariates, are calculated using site-specific intervention group codes of +.5 for intervention and -.5 for control as follows:

$$E(y_{\text{intervention}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3(0.5) + \beta_4(0.5 * \text{Time}_1) + \beta_5(0.5 * \text{Time}_2) + Xy$$

and

$$E(y_{\text{control}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3(-0.5) + \beta_4(-0.5 * \text{Time}_1) + \beta_5(-0.5 * \text{Time}_2) + Xy$$

where X represents the collection of covariates (gender, ..., maternal education) and y is their set of β s. Noting that $\text{Time}_2 = 0$ for spring pre-kindergarten scores, the difference between intervention and control group means then would be

$$E(y_{\text{intervention}}) - E(y_{\text{control}}) = \left\{ \begin{array}{l} \beta_3 + \beta_4 \text{Time}_1, \text{ for Pre - K Spring scores} \\ \beta_3 + \beta_4 \text{Time}_1 + \beta_5 \text{Time}_2, \text{ for K Spring scores} \end{array} \right\}$$

Estimates for the slope for the pre-kindergarten portion of the linear spline model are constructed through linear combinations (C) of the fixed parameters:

$$(C' \beta)_{\text{slope,Pre-K,intervention}} = \beta_1 + \beta_4(0.5)$$

$$(C' \beta)_{\text{slope,Pre-K,control}} = \beta_1 + \beta_4(-0.5)$$

The difference between intervention and control group slopes for this portion would be

$$(C' \beta)_{\text{slope,Pre-K,intervention}} - (C' \beta)_{\text{slope,Pre-K,control}} = \beta_4$$

Similarly, estimates for the slope for the kindergarten portion of the linear spline model are

$$(C' \beta)_{\text{slope,K,intervention}} = \beta_1 + \beta_2 + \beta_4(0.5) + \beta_5(0.5)$$

$$(C' \beta)_{\text{slope,K,control}} = \beta_1 + \beta_2 + \beta_4(-0.5) + \beta_5(-0.5)$$

Again, the difference between intervention and control group slopes for the kindergarten portion would be

$$(C' \beta)_{\text{slope,K,intervention}} - (C' \beta)_{\text{slope,K,control}} = \beta_4 + \beta_5$$

Example for a double intervention at a single site

A more complicated example of testing intervention differences using the linear spline model for child outcomes is described here for the University of Texas-Houston research team that implemented two intervention curricula (*Doors to Discovery* and *Lets Begin with the Letter People*) at one site (University of Texas Health-Houston) using one shared control group. The model is as follows:

$$E(y) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3(\text{LETP-TX}) + \beta_4(\text{LETP-TX} \times \text{Time}_1) + \beta_5(\text{LETP-TX} \times \text{Time}_2) + \beta_6(\text{DOOR-TX}) + \beta_7(\text{DOOR-TX} \times \text{Time}_1) + \beta_8(\text{DOOR-TX} \times \text{Time}_2) + \beta_9 \text{Gender} + \beta_{10} \text{Age} + \beta_{11} \text{Race} + \beta_{12} \text{Disability} + \beta_{13} \text{MaternalEducation}$$

Note that there are two interventions and one control group at this site. With the appropriate coding scheme, the *Let's Begin with Letter People* (LETP) intervention and shared control group means, adjusted for covariates, are as follows:

$$E(y_{\text{intervention-LETP-TX}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3 (1.0) \\ + \beta_4 (1.0 \times \text{Time}_1) + \beta_5 (1.0 \times \text{Time}_2) + X \gamma$$

and

$$E(y_{\text{control}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3 (-1.0) \\ + \beta_4 (-1.0 \times \text{Time}_1) + \beta_5 (-1.0 \times \text{Time}_2) \\ + \beta_6 (-1.0) + \beta_7 (-1.0 \times \text{Time}_1) + \beta_8 (-1.0 \times \text{Time}_2) + X \gamma$$

Noting that $\text{Time}_2 = 0$ for spring pre-kindergarten scores, the difference between the LETP intervention and shared control group means then would be:

$$E(y_{\text{intervention-LETP-TX}}) - E(y_{\text{control}}) = \begin{cases} \beta_3 (2.0) + \beta_4 (2.0 \times \text{Time}_1) + \\ \beta_6 (1.0) + \beta_7 (1.0 \times \text{Time}_1), \text{ for Pre - K Spring scores} \\ \beta_3 (2.0) + \beta_4 (2.0 \times \text{Time}_1) + \beta_5 (2.0 \times \text{Time}_2) + \\ \beta_6 (1.0) + \beta_7 (1.0 \times \text{Time}_1) + \beta_8 (1.0 \times \text{Time}_2), \text{ for K Spring scores} \end{cases}$$

Estimates for the slope for the pre-kindergarten portion of the linear spline model are

$$(C' \beta)_{\text{slope,Pre-K,intervention-LETP-TX}} = \beta_1 + \beta_4 (1.0) \\ (C' \beta)_{\text{slope,Pre-K,control}} = \beta_1 + \beta_4 (-1.0) + \beta_7 (-1.0)$$

The difference between LETP intervention and control group slopes for this portion would be

$$(C' \beta)_{\text{slope,Pre-K,intervention-LETP-TX}} - (C' \beta)_{\text{slope,Pre-K,control}} = \beta_4 (2.0) + \beta_7$$

Similarly, estimates for the slope for the kindergarten portion of the linear spline model are

$$(C' \beta)_{\text{slope,K,intervention-LETP-TX}} = \beta_1 + \beta_2 + \beta_4 (1.0) + \beta_5 (1.0) \\ (C' \beta)_{\text{slope,K,control}} = \beta_1 + \beta_2 + \beta_4 (-1.0) + \beta_5 (-1.0) + \beta_7 (-1.0) + \beta_8 (-1.0)$$

Again, the difference between LETP intervention and control group slopes for the kindergarten portion would be

$$(C' \beta)_{\text{slope,K,intervention-LETP-TX}} - (C' \beta)_{\text{slope,K,control}} = \beta_4 (2.0) + \beta_5 (2.0) + \beta_7 + \beta_8$$

Other Modifications

An example for a research team working at more than one site is a relatively straightforward modification: simply add covariates and covariate by time interactions for each site. For a research team working at two sites using one curriculum (e.g., the University of North Carolina working in both North Carolina and Georgia with *Creative Curriculum*), the following model would apply:

$$\begin{aligned}
E(y) = & \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \\
& \beta_3 (\text{CRT_NC}) + \beta_4 (\text{CRT_NC} \times \text{Time}_1) + \beta_5 (\text{CRT_NC} \times \text{Time}_2) + \\
& \beta_6 (\text{CRT_GA}) + \beta_7 (\text{CRT_GA} \times \text{Time}_1) + \beta_8 (\text{CRT_GA} \times \text{Time}_2) + \\
& \beta_9 \text{Gender} + \beta_{10} \text{Age} + \beta_{11} \text{Race} + \beta_{12} \text{Disability} + \beta_{13} \text{MaternalEducation}
\end{aligned}$$

For a research team using three sites (e.g., the Success for All Foundation) the above model is used with the addition of a line for a third site.

The above model can be used to illustrate the modifications made for the other two types of models (simple repeated measures and ANOVA). For the repeated measures model, drop the Time_2 terms creating the modified model:

$$\begin{aligned}
E(y) = & \beta_0 + \beta_1 \text{Time}_1 \\
& \beta_2 (\text{CRT_NC}) + \beta_3 (\text{CRT_NC} \times \text{Time}_1) + \\
& \beta_4 (\text{CRT_GA}) + \beta_5 (\text{CRT_GA} \times \text{Time}_1) + \\
& \beta_6 \text{Gender} + \beta_7 \text{Age} + \beta_8 \text{Race} + \beta_9 \text{Disability} + \beta_{10} \text{MaternalEducation}
\end{aligned}$$

For the ANCOVA model further drop the Time_1 terms and add the baseline measure of y , creating the modified model:

$$\begin{aligned}
E(y) = & \beta_0 + \beta_1 (y_{\text{baseline}}) \\
& \beta_2 (\text{CRT_NC}) + \\
& \beta_3 (\text{CRT_GA}) + \\
& \beta_4 \text{Gender} + \beta_5 \text{Age} + \beta_6 \text{Race} + \beta_7 \text{Disability} + \beta_8 \text{MaternalEducation}
\end{aligned}$$

Specific Time Points Used in Tests of Impact

As indicated above, the time points of interest for testing of the intervention and control group differences were start of treatment, fall pre-kindergarten, spring pre-kindergarten, and kindergarten assessments. The time points used in impact testing were mean observed values for these assessments. More specifically, the mean values across all the classrooms (treatment and control classrooms) that were used in the model contrasts, estimating group means, and slope and impacts at various points, for Time_1 were as follows: 0.0 for start of treatment; 6.608 weeks for fall pre-kindergarten assessment; 35.376 weeks for spring pre-kindergarten assessment (Time_1); and 87.454 weeks for spring kindergarten assessment. Corresponding Time_2 values used are 0's (zeros) for the pre-kindergarten year and 52.078 for kindergarten spring estimates. The estimated trajectories whose coefficients were used in the test contrasts were estimated by modeling the outcomes using Time_1 and Time_2 values derived from the *actual* specific classroom assessment (start of assessments time values averaged across all classrooms) and start of intervention dates, which varied from classroom to classroom.

Method of Estimating Degrees of Freedom

The use of normal and chi-squared distributions in testing ML estimates of regression coefficients or contrasts is generally too liberal when sample sizes are small. To use the t or F distributions requires specifying denominator degrees of freedom, which is not easy to determine in unbalanced data. To accommodate this, several approaches have been developed to approximate the denominator degrees of freedom. Options include the containment, residual, Satterthwaite, and Kenward-Roger methods. Researchers

(Kenward and Roger 1997, Keselman et al. 1998, Schaalje, McBride, and Fellingham 2002, and Gomez, Schaalje, and Fellingham 2005) have indicated that the Kenward-Roger adjustment provides the most unbiased Type I error rate for complicated covariance structures and small sample sizes. Disparities between the use of different methodologies is generally small. For these analyses, all tests of fixed effects and contrasts (i.e., intervention and control group means at various time points, estimated impacts, various slopes, and impacts on slopes) were tested using the Kenward-Roger method for estimating denominator degrees of freedom.

**Appendix C:
Unadjusted Mean Scores**

Page intentionally left blank.

Table C-1a. Unadjusted mean scores of child-level outcome measures, *Bright Beginnings*: Tennessee

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	93.07	17.43	103	99.77	14.68	98	102.58	12.77	100
Control	93.12	16.21	105	96.48	16.69	100	99.88	16.18	104
CMA-A Mathematics Composite									
Treatment	0.34	0.25	103	0.56	0.25	98	0.69	0.17	100
Control	0.34	0.24	105	0.53	0.27	100	0.69	0.18	103
Shape Composition ¹									
Treatment	0.94	0.86	103	1.79	0.93	98	2.46	0.72	100
Control	0.83	0.74	105	1.85	0.91	100	2.36	0.89	104
Reading									
TERA									
Treatment	86.09	11.95	103	92.82	15.91	98	93.30	16.02	100
Control	84.34	10.18	105	87.98	14.71	100	93.99	17.75	103
WJ Letter Word Identification									
Treatment	93.57	14.66	103	102.03	14.97	98	105.93	10.67	100
Control	88.79	13.11	105	97.21	13.03	100	103.96	13.41	104
WJ Spelling									
Treatment	89.37	13.30	103	94.21	12.46	98	101.50	12.09	100
Control	88.44	12.46	105	90.94	12.98	100	100.57	15.15	104
Phonological awareness									
Pre-CTOPPP									
Treatment	8.02	4.20	103	10.26	4.50	98	†	†	†
Control	7.55	4.25	105	10.38	4.78	100	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.26	2.76	100
Control	†	†	†	†	†	†	4.30	3.27	103
Language									
PPVT									
Treatment	90.72	16.00	103	96.55	14.71	96	98.52	10.83	100
Control	90.83	16.43	105	93.93	15.37	99	97.21	13.74	103
TOLD									
Treatment	8.16	2.45	102	9.54	2.95	97	10.54	2.91	100
Control	8.28	2.69	105	9.11	2.73	100	9.91	2.93	103
Behavior									
SSRS Social Skills									
Treatment	92.25	13.74	102	105.24	14.38	92	99.49	15.11	95
Control	97.94	25.30	104	111.94	14.39	99	102.36	17.22	96
SSRS Problem Behavior									
Treatment	103.67	12.71	102	100.45	13.54	92	102.73	13.29	95
Control	104.89	16.34	104	97.62	11.79	99	99.40	13.64	96
PLBS									
Treatment	48.20	10.61	102	53.32	11.30	91	†	†	†
Control	46.63	18.06	104	53.84	9.20	98	†	†	†
LBS									
Treatment	†	†	†	†	†	†	43.61	12.49	94
Control	†	†	†	†	†	†	47.83	11.62	96

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-1b. Unadjusted mean scores of classroom-level outcome measures, *Bright Beginnings*: Tennessee

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	4.03	0.59	7	4.18	0.95	7
Control	3.05	0.51	7	3.54	0.69	7
Teacher-child interaction						
Arnett Detachment						
Treatment	1.25	0.38	7	2.04	0.65	7
Control	2.18	0.67	7	2.18	0.83	7
Arnett Harshness						
Treatment	1.22	0.21	7	1.33	0.31	7
Control	1.49	0.53	7	1.38	0.25	7
Arnett Permissiveness						
Treatment	1.95	0.23	7	2.19	0.33	7
Control	2.14	0.47	7	2.19	0.33	7
Arnett Positive Interactions						
Treatment	3.21	0.56	7	3.17	0.75	7
Control	2.53	0.88	7	2.77	0.62	7
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	5.06	2.64	7
Control	†	†	†	2.92	2.00	7
TBRS Oral Language						
Treatment	†	†	†	3.76	1.42	7
Control	†	†	†	3.33	1.06	7
TBRS Phonological Awareness						
Treatment	†	†	†	5.00	4.04	7
Control	†	†	†	1.57	1.13	7
TBRS Print and Letter Knowledge						
Treatment	†	†	†	2.95	1.39	7
Control	†	†	†	1.71	0.40	7
TBRS Written Expression						
Treatment	†	†	†	3.48	1.23	7
Control	†	†	†	2.19	0.79	7
TBRS Math Concepts						
Treatment	†	†	†	2.78	1.36	7
Control	†	†	†	2.20	0.79	7

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-2a. Unadjusted mean scores of child-level outcome measures, *Creative Curriculum: Tennessee*

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	95.21	13.18	101	98.63	12.03	93	103.44	9.60	96
Control	93.12	16.21	105	96.48	16.69	100	99.88	16.18	104
CMA-A Mathematics Composite									
Treatment	0.37	0.22	101	0.56	0.23	93	0.70	0.17	96
Control	0.34	0.24	105	0.53	0.27	100	0.69	0.18	103
Shape Composition ¹									
Treatment	1.02	0.85	101	1.81	0.95	93	2.41	0.70	96
Control	0.83	0.74	105	1.85	0.91	100	2.36	0.89	104
Reading									
TERA									
Treatment	84.32	10.42	101	87.92	12.06	93	93.49	15.33	96
Control	84.34	10.18	105	87.98	14.71	100	93.99	17.75	103
WJ Letter Word Identification									
Treatment	93.89	13.68	101	99.32	11.06	93	107.42	11.92	96
Control	88.79	13.11	105	97.21	13.03	100	103.96	13.41	104
WJ Spelling									
Treatment	88.38	11.80	101	93.39	11.07	93	104.15	11.62	96
Control	88.44	12.46	105	90.94	12.98	100	100.57	15.15	104
Phonological awareness									
Pre-CTOPPP									
Treatment	7.61	3.70	101	10.56	3.60	93	†	†	†
Control	7.55	4.25	105	10.38	4.78	100	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.47	3.41	96
Control	†	†	†	†	†	†	4.30	3.27	103
Language									
PPVT									
Treatment	91.47	16.32	101	97.92	13.27	93	99.20	10.82	96
Control	90.83	16.43	105	93.93	15.37	99	97.21	13.74	103
TOLD									
Treatment	8.32	2.27	98	9.43	2.55	93	10.32	2.29	96
Control	8.28	2.69	105	9.11	2.73	100	9.91	2.93	103
Behavior									
SSRS Social Skills									
Treatment	95.22	18.30	101	110.23	16.53	88	106.75	14.30	84
Control	97.94	25.30	104	111.94	14.39	99	102.36	17.22	96
SSRS Problem Behavior									
Treatment	103.40	14.15	101	99.34	13.57	88	98.39	12.38	85
Control	104.89	16.34	104	97.62	11.79	99	99.40	13.64	96
PLBS									
Treatment	48.68	11.62	101	54.25	10.48	88	†	†	†
Control	46.63	18.06	104	53.84	9.20	98	†	†	†
LBS									
Treatment	†	†	†	†	†	†	49.06	9.41	84
Control	†	†	†	†	†	†	47.83	11.62	96

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-2b. Unadjusted mean scores of classroom-level outcome measures, *Creative Curriculum*: Tennessee

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	4.48	0.81	7	4.20	0.74	7
Control	3.05	0.51	7	3.54	0.69	7
Teacher-child interaction						
Arnett Detachment						
Treatment	1.64	0.92	7	1.80	0.48	5
Control	2.18	0.67	7	2.18	0.83	7
Arnett Harshness						
Treatment	1.40	0.44	7	1.29	0.33	5
Control	1.49	0.53	7	1.38	0.25	7
Arnett Permissiveness						
Treatment	2.24	0.32	7	2.33	0.24	5
Control	2.14	0.47	7	2.19	0.33	7
Arnett Positive Interactions						
Treatment	2.99	0.89	7	2.96	0.34	5
Control	2.53	0.88	7	2.77	0.62	7
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	3.26	2.12	5
Control	†	†	†	2.92	2.00	7
TBRS Oral Language						
Treatment	†	†	†	3.83	1.72	5
Control	†	†	†	3.33	1.06	7
TBRS Phonological Awareness						
Treatment	†	†	†	4.00	2.00	5
Control	†	†	†	1.57	1.13	7
TBRS Print and Letter Knowledge						
Treatment	†	†	†	2.50	0.51	5
Control	†	†	†	1.71	0.40	7
TBRS Written Expression						
Treatment	†	†	†	3.27	0.55	5
Control	†	†	†	2.19	0.79	7
TBRS Math Concepts						
Treatment	†	†	†	3.03	0.58	5
Control	†	†	†	2.20	0.79	7

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-3a. Unadjusted mean scores of child-level outcome measures, *Creative Curriculum*: North Carolina and Georgia

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	90.17	14.82	94	91.36	12.26	89	95.54	14.29	85
Control	86.87	16.97	95	89.45	13.75	80	93.46	13.21	76
CMA-A Mathematics Composite									
Treatment	0.29	0.22	94	0.41	0.27	90	0.68	0.18	85
Control	0.29	0.20	95	0.44	0.29	80	0.63	0.20	76
Shape Composition ¹									
Treatment	.85	0.91	93	1.43	0.89	89	2.08	0.80	85
Control	.82	0.89	95	1.25	0.83	80	2.05	0.92	76
Reading									
TERA									
Treatment	83.73	11.76	95	83.74	13.97	90	93.28	17.62	85
Control	81.81	10.90	95	86.39	13.88	80	92.51	15.30	76
WJ Letter Word Identification									
Treatment	95.53	17.58	94	99.15	12.11	89	105.78	15.25	85
Control	96.80	18.42	95	101.74	13.08	80	105.28	12.95	76
WJ Spelling									
Treatment	90.39	13.26	94	88.93	14.38	89	102.93	17.90	85
Control	89.96	13.92	95	91.95	13.23	80	102.28	16.25	76
Phonological awareness									
Pre-CTOPPP									
Treatment	6.67	3.49	95	8.44	4.08	90	†	†	†
Control	6.49	2.90	95	8.19	4.03	81	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	2.81	3.03	85
Control	†	†	†	†	†	†	2.51	2.83	76
Language									
PPVT									
Treatment	81.49	14.89	95	86.02	14.43	87	90.98	11.94	84
Control	81.03	15.30	94	85.42	13.40	78	88.09	13.60	76
TOLD									
Treatment	7.91	2.24	93	7.97	2.58	89	9.22	2.67	85
Control	7.51	2.41	94	8.44	2.68	80	9.63	2.88	76
Behavior									
SSRS Social Skills									
Treatment	104.21	14.43	96	106.16	14.21	88	96.02	14.24	60
Control	101.27	14.77	95	104.40	10.79	81	96.80	13.22	44
SSRS Problem Behavior									
Treatment	98.90	12.50	96	99.06	13.08	89	104.17	14.81	59
Control	100.61	14.68	95	102.37	14.26	81	102.49	12.78	45
PLBS									
Treatment	51.21	9.65	97	52.29	9.58	89	†	†	†
Control	48.74	10.69	95	50.52	10.87	81	†	†	†
LBS									
Treatment	†	†	†	†	†	†	44.15	12.43	59
Control	†	†	†	†	†	†	45.53	11.10	45

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-3b. Unadjusted mean scores of classroom-level outcome measures, *Creative Curriculum*: North Carolina and Georgia

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	3.64	0.88	9	4.36	0.50	9
Control	3.41	1.27	9	3.32	0.94	9
Teacher-child interaction						
Arnett Detachment						
Treatment	1.56	0.72	8	1.19	0.27	9
Control	2.00	0.82	9	2.31	0.96	9
Arnett Harshness						
Treatment	1.71	0.81	8	1.42	0.43	9
Control	2.14	1.07	9	1.83	0.68	9
Arnett Permissiveness						
Treatment	2.33	0.76	8	1.74	0.28	9
Control	2.00	0.58	9	2.30	0.61	9
Arnett Positive Interactions						
Treatment	2.54	0.77	8	3.20	0.70	9
Control	2.28	0.82	9	2.14	0.68	9
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	1.87	2.25	9
Control	†	†	†	1.86	1.80	8
TBRS Oral Language						
Treatment	†	†	†	4.03	1.34	9
Control	†	†	†	2.04	0.93	8
TBRS Phonological Awareness						
Treatment	†	†	†	4.22	4.06	9
Control	†	†	†	3.75	1.98	8
TBRS Print and Letter Knowledge						
Treatment	†	†	†	3.24	1.39	9
Control	†	†	†	2.13	0.49	8
TBRS Written Expression						
Treatment	†	†	†	3.78	1.29	9
Control	†	†	†	1.83	0.59	8
TBRS Math Concepts						
Treatment	†	†	†	2.83	1.99	9
Control	†	†	†	1.39	0.42	8

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-4a. Unadjusted mean scores of child-level outcome measures, *Creative Curriculum with Ladders to Literacy*: New Hampshire

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	92.33	15.56	61	94.20	14.25	54	95.46	14.70	46
Control	93.78	11.89	59	96.10	15.39	51	101.26	9.12	34
CMA-A Mathematics Composite									
Treatment	0.45	0.26	62	0.59	0.23	54	0.64	0.20	46
Control	0.42	0.23	60	0.56	0.25	51	0.71	0.13	35
Shape Composition ¹									
Treatment	1.36	0.88	61	1.75	1.00	53	2.48	0.78	46
Control	1.43	0.81	60	1.75	0.98	51	2.63	0.69	35
Reading									
TERA									
Treatment	84.76	10.85	62	82.76	14.66	54	85.76	16.99	46
Control	82.62	11.65	61	85.27	14.66	51	92.03	12.48	35
WJ Letter Word Identification									
Treatment	94.26	12.91	61	96.30	13.20	54	96.11	12.75	46
Control	93.95	13.33	60	97.90	13.56	51	100.60	13.92	35
WJ Spelling									
Treatment	86.66	12.32	61	92.04	12.13	54	91.54	16.18	46
Control	89.60	11.42	60	89.96	15.12	51	98.29	12.92	35
Phonological awareness									
Pre-CTOPPP									
Treatment	7.61	3.60	62	8.83	4.60	54	†	†	†
Control	7.79	4.71	61	9.10	5.09	51	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.02	3.66	46
Control	†	†	†	†	†	†	4.60	4.55	35
Language									
PPVT									
Treatment	89.79	16.68	62	90.80	18.03	54	94.36	13.33	45
Control	90.46	14.99	61	95.43	14.88	51	100.23	9.24	35
TOLD									
Treatment	8.39	2.53	62	9.11	2.87	53	9.20	2.58	46
Control	8.25	2.55	61	9.45	2.61	51	9.74	1.93	35
Behavior									
SSRS Social Skills									
Treatment	96.11	19.73	62	101.37	15.61	51	98.72	13.77	32
Control	102.47	15.07	60	105.78	12.83	49	96.60	10.39	15
SSRS Problem Behavior									
Treatment	103.32	13.83	62	103.27	13.78	51	103.46	13.13	35
Control	104.07	15.36	60	104.35	13.70	49	106.20	13.03	15
PLBS									
Treatment	49.03	11.15	62	49.61	12.11	51	†	†	†
Control	51.55	9.01	60	51.22	8.24	49	†	†	†
LBS									
Treatment	†	†	†	†	†	†	44.11	10.35	35
Control	†	†	†	†	†	†	45.47	9.83	15

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-4b. Unadjusted mean scores of classroom-level outcome measures, *Creative Curriculum with Ladders to Literacy*: New Hampshire

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	4.56	1.04	7	4.04	0.92	7
Control	3.93	0.40	7	4.23	0.67	7
Teacher-child interaction						
Arnett Detachment						
Treatment	1.43	0.66	7	1.93	1.06	7
Control	1.50	0.54	7	1.50	0.67	6
Arnett Harshness						
Treatment	1.54	0.33	7	1.33	0.41	7
Control	1.37	0.17	7	1.37	0.40	6
Arnett Permissiveness						
Treatment	2.00	0.33	7	2.38	0.52	7
Control	1.90	0.16	7	2.11	0.17	6
Arnett Positive Interactions						
Treatment	3.43	0.79	7	2.74	1.01	7
Control	3.10	0.80	7	2.90	0.84	6
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	2.43	2.28	7
Control	†	†	†	3.48	2.35	6
TBRS Oral Language						
Treatment	†	†	†	3.14	1.30	7
Control	†	†	†	3.98	1.01	6
TBRS Phonological Awareness						
Treatment	†	†	†	3.14	2.79	7
Control	†	†	†	4.17	3.37	6
TBRS Print and Letter Knowledge						
Treatment	†	†	†	2.26	0.71	7
Control	†	†	†	2.06	1.19	6
TBRS Written Expression						
Treatment	†	†	†	2.83	1.13	7
Control	†	†	†	2.22	1.07	6
TBRS Math Concepts						
Treatment	†	†	†	2.06	1.06	7
Control	†	†	†	2.29	1.81	6

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-5a. Unadjusted mean scores of child-level outcome measures, Curiosity Corner: Florida, Kansas, and New Jersey

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	90.14	14.99	100	93.74	13.72	93	98.76	12.71	95
Control	95.11	14.36	94	96.95	15.97	87	100.91	12.40	93
CMA-A Mathematics Composite									
Treatment	0.46	0.24	105	0.63	0.23	101	0.68	0.17	97
Control	0.50	0.24	106	0.66	0.23	103	0.70	0.17	97
Shape Composition ¹									
Treatment	1.20	0.94	105	1.74	1.00	101	2.41	0.83	97
Control	1.31	1.03	106	1.67	0.98	99	2.22	0.93	97
Reading									
TERA									
Treatment	86.54	11.76	105	87.45	13.99	101	91.40	15.18	95
Control	86.60	11.59	99	89.30	14.46	99	89.85	14.55	93
WJ Letter Word Identification									
Treatment	97.40	14.92	102	98.39	13.11	88	107.98	13.67	96
Control	95.90	15.50	102	100.91	14.14	89	106.18	12.44	93
WJ Spelling									
Treatment	93.70	18.41	102	97.21	15.31	96	106.51	15.79	91
Control	97.47	11.65	99	100.17	14.17	98	106.03	11.53	91
Phonological awareness									
Pre-CTOPPP									
Treatment	6.54	4.23	105	9.44	4.34	101	†	†	†
Control	7.56	3.43	104	10.27	4.44	103	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.10	3.77	97
Control	†	†	†	†	†	†	3.89	3.88	96
Language									
PPVT									
Treatment	85.83	14.83	105	90.44	14.30	101	94.53	12.56	96
Control	90.61	14.91	101	93.76	16.61	100	95.44	13.53	93
TOLD									
Treatment	8.04	2.56	104	8.50	2.70	100	9.87	2.73	95
Control	8.67	2.49	99	9.14	2.72	99	10.02	2.45	95
Behavior									
SSRS Social Skills									
Treatment	96.11	13.84	101	104.20	12.46	100	100.46	16.14	81
Control	103.90	14.27	102	108.56	13.86	101	99.28	14.81	88
SSRS Problem Behavior									
Treatment	103.67	13.27	101	102.39	12.34	100	100.99	15.22	82
Control	97.62	11.68	103	97.80	12.05	100	99.49	12.99	88
PLBS									
Treatment	47.95	11.20	100	49.26	11.57	99	†	†	†
Control	53.13	9.67	106	52.75	9.95	104	†	†	†
LBS									
Treatment	†	†	†	†	†	†	46.05	11.96	84
Control	†	†	†	†	†	†	47.32	8.33	92

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-5b. Unadjusted mean scores of classroom-level outcome measures, *Curiosity Corner*: Florida, Kansas, and New Jersey

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	4.95	0.86	14	4.72	0.78	14
Control	4.97	0.62	17	4.79	0.63	17
Teacher-child interaction						
Arnett Detachment						
Treatment	1.08	0.21	13	1.32	0.54	14
Control	1.06	0.24	17	1.22	0.34	16
Arnett Harshness						
Treatment	1.50	0.57	13	1.51	0.65	14
Control	1.53	0.26	17	1.49	0.33	16
Arnett Permissiveness						
Treatment	3.21	0.79	13	3.55	0.70	14
Control	3.39	0.34	17	3.81	0.36	16
Arnett Positive Interactions						
Treatment	3.56	0.59	13	3.48	0.50	14
Control	3.16	0.48	17	3.53	0.43	16
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	5.67	3.00	14
Control	†	†	†	2.49	1.50	17
TBRS Oral Language						
Treatment	†	†	†	4.42	0.96	14
Control	†	†	†	4.00	1.15	17
TBRS Phonological Awareness						
Treatment	†	†	†	5.93	2.43	14
Control	†	†	†	5.24	2.63	17
TBRS Print and Letter Knowledge						
Treatment	†	†	†	2.96	0.98	14
Control	†	†	†	2.80	1.11	17
TBRS Written Expression						
Treatment	†	†	†	2.50	0.76	14
Control	†	†	†	1.94	1.00	17
TBRS Math Concepts						
Treatment	†	†	†	2.18	0.95	14
Control	†	†	†	2.50	1.27	17

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-6a. Unadjusted mean scores of child-level outcome measures, Doors to Discovery: Texas

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	96.90	16.96	100	100.30	13.24	94	102.63	10.95	83
Control	95.70	17.60	94	99.28	16.60	89	102.40	11.38	72
CMA-A Mathematics Composite									
Treatment	0.44	0.24	100	0.70	0.20	94	0.70	0.16	83
Control	0.44	0.27	94	0.65	0.24	89	0.72	0.14	72
Shape Composition ¹									
Treatment	1.36	0.88	100	1.64	0.84	94	2.46	0.79	83
Control	1.44	1.00	94	1.72	0.69	89	2.51	0.69	72
Reading									
TERA									
Treatment	89.65	15.93	100	93.78	17.22	94	93.57	18.88	83
Control	85.91	15.56	94	92.76	17.86	88	93.96	16.47	72
WJ Letter Word Identification									
Treatment	101.87	16.33	100	107.21	14.56	94	108.37	14.82	83
Control	100.46	17.21	94	106.04	13.82	89	109.53	13.57	72
WJ Spelling									
Treatment	93.96	13.56	100	99.12	12.56	94	102.89	14.84	83
Control	92.30	13.63	94	97.37	12.63	89	103.46	13.14	72
Phonological awareness									
Pre-CTOPPP									
Treatment	7.74	4.79	100	11.56	4.18	93	†	†	†
Control	6.97	4.84	94	10.11	4.64	89	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	5.13	3.84	83
Control	†	†	†	†	†	†	5.04	4.24	72
Language									
PPVT									
Treatment	88.52	21.52	100	96.02	18.20	94	97.23	17.46	81
Control	83.83	20.25	94	91.33	18.12	89	94.00	16.01	71
TOLD									
Treatment	9.55	2.95	100	10.07	3.06	94	10.41	3.19	83
Control	8.40	2.82	94	9.33	2.71	89	10.08	2.80	72
Behavior									
SSRS Social Skills									
Treatment	104.38	16.03	95	108.35	16.25	92	100.99	13.73	67
Control	104.98	15.44	88	109.26	12.88	70	102.67	13.24	42
SSRS Problem Behavior									
Treatment	95.89	12.31	99	98.47	14.38	91	99.88	13.13	64
Control	100.78	14.43	87	101.61	13.67	70	97.38	11.95	42
PLBS									
Treatment	53.46	9.30	99	54.01	9.44	92	†	†	†
Control	52.53	9.73	86	54.46	8.43	69	†	†	†
LBS									
Treatment	†	†	†	†	†	†	46.19	8.24	67
Control	†	†	†	†	†	†	48.46	7.17	41

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-6b. Unadjusted mean scores of classroom-level outcome measures, *Doors to Discovery: Texas*

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	3.42	0.48	13	3.69	0.60	13
Control	3.29	0.50	16	3.45	0.83	16
Teacher-child interaction						
Arnett Detachment						
Treatment	1.48	0.51	13	1.63	0.63	12
Control	1.70	0.77	14	1.63	0.75	15
Arnett Harshness						
Treatment	1.45	0.36	13	1.76	0.79	12
Control	1.67	0.68	14	2.04	0.85	15
Arnett Permissiveness						
Treatment	2.03	0.35	13	2.11	0.36	12
Control	1.69	0.33	14	2.11	0.21	15
Arnett Positive Interactions						
Treatment	2.90	0.71	13	3.16	0.75	12
Control	2.70	0.82	14	2.91	0.85	15
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	5.71	2.72	12
Control	†	†	†	3.79	2.33	15
TBRS Oral Language						
Treatment	†	†	†	3.88	1.73	12
Control	†	†	†	3.16	1.39	15
TBRS Phonological Awareness						
Treatment	†	†	†	6.08	3.40	12
Control	†	†	†	4.00	3.66	15
TBRS Print and Letter Knowledge						
Treatment	†	†	†	3.40	1.26	12
Control	†	†	†	2.62	1.06	15
TBRS Written Expression						
Treatment	†	†	†	3.17	1.23	12
Control	†	†	†	2.60	1.03	15
TBRS Math Concepts						
Treatment	†	†	†	2.52	1.34	12
Control	†	†	†	1.97	0.88	15

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-7a. Unadjusted mean scores of child-level outcome measures, *Let's Begin with the Letter People*: Texas

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	94.55	15.94	99	98.54	13.25	95	101.42	13.42	79
Control	95.70	17.60	94	99.28	16.60	89	102.40	11.38	72
CMA-A Mathematics Composite									
Treatment	0.46	0.25	99	0.69	0.22	95	0.71	0.18	79
Control	0.44	0.27	94	0.65	0.24	89	0.72	0.14	72
Shape Composition ¹									
Treatment	1.14	1.08	99	1.93	0.95	95	2.48	0.77	79
Control	1.44	1.00	94	1.72	0.69	89	2.51	0.69	72
Reading									
TERA									
Treatment	86.19	13.97	99	93.27	16.06	95	94.76	18.50	79
Control	85.91	15.56	94	92.76	17.86	88	93.96	16.47	72
WJ Letter Word Identification									
Treatment	98.70	15.46	99	107.81	12.54	95	108.13	13.44	79
Control	100.46	17.21	94	106.04	13.82	89	109.53	13.57	72
WJ Spelling									
Treatment	92.17	10.86	99	100.53	13.01	95	103.61	15.68	79
Control	92.30	13.63	94	97.37	12.63	89	103.46	13.14	72
Phonological awareness									
Pre-CTOPPP									
Treatment	7.26	4.49	99	9.76	5.07	95	†	†	†
Control	6.97	4.84	94	10.11	4.64	89	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.72	3.66	79
Control	†	†	†	†	†	†	5.04	4.24	72
Language									
PPVT									
Treatment	87.26	20.62	99	92.27	19.18	95	95.23	18.03	79
Control	83.83	20.25	94	91.33	18.12	89	94.00	16.01	71
TOLD									
Treatment	8.82	3.00	99	9.89	2.73	95	9.99	3.12	79
Control	8.40	2.82	94	9.33	2.71	89	10.08	2.80	72
Behavior									
SSRS Social Skills									
Treatment	98.87	12.64	100	106.62	13.53	90	103.98	15.33	58
Control	104.98	15.44	88	109.26	12.88	70	102.67	13.24	42
SSRS Problem Behavior									
Treatment	99.41	12.97	100	100.26	13.24	90	97.38	11.97	58
Control	100.78	14.43	87	101.61	13.67	70	97.38	11.95	42
PLBS									
Treatment	50.30	10.89	99	51.41	11.02	90	†	†	†
Control	52.53	9.73	86	54.46	8.43	69	†	†	†
LBS									
Treatment	†	†	†	†	†	†	47.83	7.28	58
Control	†	†	†	†	†	†	48.46	7.17	41

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-7b. Unadjusted mean scores of classroom-level outcome measures, *Let's Begin with the Letter People*: Texas

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	3.32	0.63	15	4.02	0.65	15
Control	3.29	0.50	16	3.45	0.83	16
Teacher-child interaction						
Arnett Detachment						
Treatment	1.80	0.61	15	1.54	0.67	13
Control	1.70	0.77	14	1.63	0.75	15
Arnett Harshness						
Treatment	1.59	0.37	15	1.37	0.27	13
Control	1.67	0.68	14	2.04	0.85	15
Arnett Permissiveness						
Treatment	2.07	0.47	15	2.05	0.30	13
Control	1.69	0.33	14	2.11	0.21	15
Arnett Positive Interactions						
Treatment	2.77	0.65	15	3.32	0.65	13
Control	2.70	0.82	14	2.91	0.85	15
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	4.63	2.68	13
Control	†	†	†	3.79	2.33	15
TBRS Oral Language						
Treatment	†	†	†	3.65	1.50	13
Control	†	†	†	3.16	1.39	15
TBRS Phonological Awareness						
Treatment	†	†	†	6.69	4.25	13
Control	†	†	†	4.00	3.66	15
TBRS Print and Letter Knowledge						
Treatment	†	†	†	3.60	1.39	13
Control	†	†	†	2.62	1.06	15
TBRS Written Expression						
Treatment	†	†	†	3.33	1.57	13
Control	†	†	†	2.60	1.03	15
TBRS Math Concepts						
Treatment	†	†	†	2.48	1.30	13
Control	†	†	†	1.97	0.88	15

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-8a. Unadjusted mean scores of child-level outcome measures, *Early Literacy and Learning Model: Florida—University of North Florida*

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	91.55	13.35	137	93.97	10.79	124	98.64	10.93	121
Control	91.77	12.86	105	92.53	15.47	99	96.00	13.86	97
CMA-A Mathematics Composite									
Treatment	0.40	0.29	137	0.56	0.24	124	0.65	0.19	121
Control	0.41	0.26	105	0.55	0.25	100	0.66	0.21	97
Shape Composition ¹									
Treatment	1.22	0.86	137	1.34	0.88	124	2.30	0.78	121
Control	1.13	0.81	105	1.45	0.94	100	2.28	0.82	96
Reading									
TERA									
Treatment	86.87	12.18	137	91.93	13.64	124	93.42	14.06	121
Control	87.31	12.68	105	91.96	15.18	100	91.58	16.81	97
WJ Letter Word Identification									
Treatment	98.97	17.16	137	104.76	12.67	124	107.19	11.63	121
Control	102.69	16.61	105	105.97	14.14	100	108.04	14.30	97
WJ Spelling									
Treatment	91.61	11.67	137	96.46	12.97	124	104.07	13.51	120
Control	96.34	15.74	105	96.23	17.47	99	104.89	16.74	97
Phonological awareness									
Pre-CTOPPP									
Treatment	7.11	3.08	137	9.82	3.79	124	†	†	†
Control	7.54	3.29	105	9.41	3.79	100	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.10	3.29	121
Control	†	†	†	†	†	†	3.87	3.30	97
Language									
PPVT									
Treatment	85.52	13.03	137	89.36	11.30	124	92.60	12.05	121
Control	84.30	13.80	105	88.91	14.47	100	91.33	13.90	97
TOLD									
Treatment	8.48	2.51	137	9.15	2.16	124	10.07	2.36	120
Control	8.59	2.18	105	8.96	2.47	99	9.14	2.81	97
Behavior									
SSRS Social Skills									
Treatment	96.25	17.22	129	102.55	16.46	120	100.94	16.95	78
Control	99.08	12.60	103	101.32	16.02	96	97.18	18.45	76
SSRS Problem Behavior									
Treatment	96.56	12.78	130	98.50	13.82	120	101.18	15.05	77
Control	102.46	12.21	103	103.04	13.94	96	101.99	14.98	76
PLBS									
Treatment	49.95	10.68	130	51.44	10.74	120	†	†	†
Control	48.13	9.97	104	48.71	10.68	96	†	†	†
LBS									
Treatment	†	†	†	†	†	†	45.96	10.71	76
Control	†	†	†	†	†	†	44.48	13.59	75

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-8b. Unadjusted mean scores of classroom-level outcome measures, *Early Literacy and Learning Model*: Florida—University of North Florida

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	2.97	0.79	14	2.97	0.81	14
Control	3.13	1.31	14	2.97	1.39	14
Teacher-child interaction						
Arnett Detachment						
Treatment	1.73	0.67	14	2.00	0.71	14
Control	1.70	1.02	14	2.17	0.93	12
Arnett Harshness						
Treatment	2.09	0.82	14	2.14	0.87	14
Control	2.02	0.98	14	2.33	1.00	12
Arnett Permissiveness						
Treatment	2.10	0.48	14	2.02	0.38	14
Control	2.12	0.34	14	2.17	0.48	12
Arnett Positive Interactions						
Treatment	2.45	0.74	14	2.39	0.86	14
Control	2.64	0.85	14	2.11	0.85	12
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	3.43	2.85	14
Control	†	†	†	2.15	1.96	12
TBRS Oral Language						
Treatment	†	†	†	2.45	1.28	14
Control	†	†	†	2.12	1.52	12
TBRS Phonological Awareness						
Treatment	†	†	†	3.79	2.52	14
Control	†	†	†	2.42	2.23	12
TBRS Print and Letter Knowledge						
Treatment	†	†	†	2.61	0.77	14
Control	†	†	†	1.90	0.97	12
TBRS Written Expression						
Treatment	†	†	†	1.88	0.72	14
Control	†	†	†	1.75	0.91	12
TBRS Math Concepts						
Treatment	†	†	†	1.12	0.48	14
Control	†	†	†	1.38	0.95	12

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-9a. Unadjusted mean scores of child-level outcome measures, *Language-Focused Curriculum*: Virginia

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	96.87	12.34	82	98.89	9.63	92	100.08	10.11	92
Control	97.48	12.57	83	96.57	11.53	94	98.41	10.42	93
CMA-A Mathematics Composite									
Treatment	0.42	0.22	89	0.63	0.24	92	0.67	0.16	94
Control	0.41	0.22	92	0.60	0.20	95	0.66	0.18	95
Shape Composition ¹									
Treatment	0.98	0.75	89	1.60	0.77	92	2.44	0.82	94
Control	0.92	0.84	92	1.53	0.85	95	2.39	0.80	95
Reading									
TERA									
Treatment	88.09	9.67	86	90.81	11.69	91	92.09	12.10	94
Control	87.36	9.00	90	88.34	10.81	95	91.45	12.35	93
WJ Letter Word Identification									
Treatment	97.72	15.11	86	105.32	12.21	90	109.50	9.76	92
Control	95.99	13.26	86	103.42	12.39	95	108.16	9.67	95
WJ Spelling									
Treatment	96.30	13.73	86	99.77	14.12	81	107.85	12.68	89
Control	94.72	12.24	88	96.50	14.55	84	106.01	13.40	90
Phonological awareness									
Pre-CTOPPP									
Treatment	6.76	4.07	89	10.39	3.99	93	†	†	†
Control	6.84	3.13	93	9.60	3.97	95	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.09	2.68	94
Control	†	†	†	†	†	†	3.91	3.20	95
Language									
PPVT									
Treatment	94.81	13.42	89	99.43	12.12	93	99.78	10.43	91
Control	95.76	11.33	93	99.36	10.89	95	100.27	10.49	95
TOLD									
Treatment	8.68	2.29	88	9.59	2.16	92	10.02	2.32	93
Control	8.74	2.66	91	9.55	2.55	95	9.99	2.32	94
Behavior									
SSRS Social Skills									
Treatment	101.74	14.37	90	103.17	15.12	90	97.46	12.27	74
Control	102.53	16.73	78	110.37	16.46	91	97.39	16.67	80
SSRS Problem Behavior									
Treatment	104.19	13.01	90	100.78	12.51	90	101.14	13.19	77
Control	98.08	13.58	79	94.98	12.50	91	100.23	13.70	81
PLBS									
Treatment	50.36	8.99	78	52.06	10.02	90	†	†	†
Control	53.81	9.60	79	55.42	10.00	92	†	†	†
LBS									
Treatment	†	†	†	†	†	†	46.17	11.04	76
Control	†	†	†	†	†	†	44.73	10.76	80

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-9b. Unadjusted mean scores of classroom-level outcome measures, *Language-Focused Curriculum: Virginia*

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	—	—	—	5.02	0.66	7
Control	—	—	—	4.76	0.37	7
Teacher-child interaction						
Arnett Detachment						
Treatment	—	—	—	1.25	0.46	7
Control	—	—	—	1.63	1.07	6
Arnett Harshness						
Treatment	—	—	—	1.22	0.13	7
Control	—	—	—	1.39	0.15	6
Arnett Permissiveness						
Treatment	—	—	—	3.76	0.16	7
Control	—	—	—	3.72	0.25	6
Arnett Positive Interactions						
Treatment	—	—	—	3.39	0.67	7
Control	—	—	—	3.13	0.52	6
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	4.71	1.56	7
Control	†	†	†	4.33	2.21	7
TBRS Oral Language						
Treatment	†	†	†	5.27	1.08	7
Control	†	†	†	4.71	0.99	7
TBRS Phonological Awareness						
Treatment	†	†	†	6.86	1.77	7
Control	†	†	†	5.71	1.80	7
TBRS Print and Letter Knowledge						
Treatment	†	†	†	3.29	1.04	7
Control	†	†	†	2.88	0.76	7
TBRS Written Expression						
Treatment	†	†	†	3.05	0.95	7
Control	†	†	†	2.38	0.49	7
TBRS Math Concepts						
Treatment	†	†	†	2.63	0.76	7
Control	†	†	†	2.00	0.54	7

— Not available. Data were collected but not reported.

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-10a. Unadjusted mean scores of child-level outcome measures, *Literacy Express*: Florida—Florida State University

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	89.05	15.64	84	90.71	15.37	89	92.42	17.45	71
Control	87.49	13.10	86	87.86	13.77	88	90.54	14.23	76
CMA-A Mathematics Composite									
Treatment	0.47	0.27	89	0.55	0.24	92	0.57	0.22	75
Control	0.46	0.26	90	0.52	0.21	93	0.57	0.21	79
Shape Composition ¹									
Treatment	1.22	0.95	89	1.68	1.00	92	2.16	0.97	74
Control	1.20	0.91	89	1.55	0.94	93	2.11	0.91	79
Reading									
TERA									
Treatment	84.79	12.40	86	85.10	12.45	90	83.82	15.03	74
Control	83.97	10.74	88	81.10	11.00	93	82.06	13.38	79
WJ Letter Word Identification									
Treatment	98.61	16.01	87	100.05	14.10	91	102.01	14.35	73
Control	91.53	14.94	83	95.60	12.38	92	99.74	12.15	78
WJ Spelling									
Treatment	93.58	13.11	84	89.91	14.34	90	99.37	17.67	71
Control	91.07	12.94	86	87.67	11.75	92	97.83	13.15	76
Phonological awareness									
Pre-CTOPPP									
Treatment	6.66	3.89	91	10.08	4.54	93	†	†	†
Control	6.75	3.85	91	8.79	4.37	95	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	3.20	3.43	76
Control	†	†	†	†	†	†	2.68	2.98	79
Language									
PPVT									
Treatment	87.52	14.32	90	92.55	15.43	92	94.23	15.12	73
Control	82.83	15.67	90	87.31	13.77	94	89.23	12.46	78
TOLD									
Treatment	7.29	3.06	83	8.47	2.62	89	9.11	3.42	75
Control	7.46	2.66	82	8.33	2.71	92	8.44	3.25	79
Behavior									
SSRS Social Skills									
Treatment	102.28	16.01	76	105.58	15.86	89	94.46	14.38	76
Control	95.06	14.21	90	104.60	15.59	88	97.93	15.28	74
SSRS Problem Behavior									
Treatment	96.77	11.03	93	99.30	11.43	90	101.93	13.49	76
Control	103.57	13.57	93	104.27	14.31	88	102.73	13.62	78
PLBS									
Treatment	52.72	9.80	93	52.87	10.32	89	†	†	†
Control	47.53	10.31	93	49.35	12.25	89	†	†	†
LBS									
Treatment	†	†	†	†	†	†	43.78	11.28	76
Control	†	†	†	†	†	†	45.21	7.93	77

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-10b. Unadjusted mean scores of classroom-level outcome measures, Literacy Express: Florida—Florida State University

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	5.46	0.55	10	5.40	0.82	10
Control	5.35	0.60	9	4.60	0.76	9
Teacher-child interaction						
Arnett Detachment						
Treatment	1.13	0.27	10	1.05	0.11	10
Control	1.11	0.18	9	1.36	0.55	9
Arnett Harshness						
Treatment	1.24	0.20	10	1.26	0.07	10
Control	1.30	0.17	9	1.47	0.51	9
Arnett Permissiveness						
Treatment	3.40	0.49	10	3.63	0.25	10
Control	3.67	0.17	9	3.48	0.53	9
Arnett Positive Interactions						
Treatment	3.16	0.41	10	3.28	0.39	10
Control	3.39	0.31	9	3.06	0.63	9
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	5.09	1.67	10
Control	†	†	†	3.71	1.82	9
TBRS Oral Language						
Treatment	†	†	†	4.87	0.92	10
Control	†	†	†	4.73	0.77	9
TBRS Phonological Awareness						
Treatment	†	†	†	6.30	2.98	10
Control	†	†	†	3.33	2.40	9
TBRS Print and Letter Knowledge						
Treatment	†	†	†	3.63	0.78	10
Control	†	†	†	2.80	0.86	9
TBRS Written Expression						
Treatment	†	†	†	3.27	0.83	10
Control	†	†	†	3.22	1.14	9
TBRS Math Concepts						
Treatment	†	†	†	3.03	1.05	10
Control	†	†	†	3.16	1.37	9

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-11a. Unadjusted mean scores of child-level outcome measures, DLM Early Childhood Express supplemented with Open Court Reading Pre-K: Florida—Florida State University

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	90.88	16.18	95	94.14	14.23	97	99.05	13.42	82
Control	87.49	13.10	86	87.86	13.77	88	90.54	14.23	76
CMA-A Mathematics Composite									
Treatment	0.45	0.25	100	0.58	0.24	98	0.63	0.20	82
Control	0.46	0.26	90	0.52	0.21	93	0.57	0.21	79
Shape Composition ¹									
Treatment	1.16	0.94	100	1.82	1.01	98	2.28	0.82	82
Control	1.20	0.91	89	1.55	0.94	93	2.11	0.91	79
Reading									
TERA									
Treatment	88.97	11.84	98	89.99	14.33	97	91.84	15.64	80
Control	83.97	10.74	88	81.10	11.00	93	82.06	13.38	79
WJ Letter Word Identification									
Treatment	98.79	15.39	98	103.61	15.56	98	107.57	13.87	82
Control	91.53	14.94	83	95.60	12.38	92	99.74	12.15	78
WJ Spelling									
Treatment	95.46	14.35	96	95.10	14.36	96	102.69	17.23	81
Control	91.07	12.94	86	87.67	11.75	92	97.83	13.15	76
Phonological awareness									
Pre-CTOPPP									
Treatment	7.14	3.60	100	10.28	4.36	98	†	†	†
Control	6.75	3.85	91	8.79	4.37	95	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.00	3.53	82
Control	†	†	†	†	†	†	2.68	2.98	79
Language									
PPVT									
Treatment	89.00	13.57	100	94.12	12.67	98	96.14	10.77	79
Control	82.83	15.67	90	87.31	13.77	94	89.23	12.46	78
TOLD									
Treatment	8.62	2.79	95	9.39	2.51	95	10.05	2.71	82
Control	7.46	2.66	82	8.33	2.71	92	8.44	3.25	79
Behavior									
SSRS Social Skills									
Treatment	99.08	14.60	85	102.22	15.23	97	96.83	15.68	81
Control	95.06	14.21	90	104.60	15.59	88	97.93	15.28	74
SSRS Problem Behavior									
Treatment	103.75	14.13	97	106.25	15.56	96	103.07	16.05	81
Control	103.57	13.57	93	104.27	14.31	88	102.73	13.62	78
PLBS									
Treatment	46.84	12.88	96	47.99	10.84	97	†	†	†
Control	47.53	10.31	93	49.35	12.25	89	†	†	†
LBS									
Treatment	†	†	†	†	†	†	44.51	11.34	81
Control	†	†	†	†	†	†	45.21	7.93	77

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-11b. Unadjusted mean scores of classroom-level outcome measures, DLM Early Childhood Express supplemented with Open Court Reading Pre-K: Florida—Florida State University

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	5.24	0.63	11	4.77	0.82	11
Control	5.35	0.60	9	4.60	0.76	9
Teacher-child interaction						
Arnett Detachment						
Treatment	1.18	0.23	11	1.39	0.64	11
Control	1.11	0.18	9	1.36	0.55	9
Arnett Harshness						
Treatment	1.25	0.11	11	1.35	0.21	11
Control	1.30	0.17	9	1.47	0.51	9
Arnett Permissiveness						
Treatment	3.52	0.35	11	3.45	0.45	11
Control	3.67	0.17	9	3.48	0.53	9
Arnett Positive Interactions						
Treatment	3.11	0.41	11	3.15	0.63	11
Control	3.39	0.31	9	3.06	0.63	9
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	3.60	2.11	11
Control	†	†	†	3.71	1.82	9
TBRS Oral Language						
Treatment	†	†	†	4.30	0.74	11
Control	†	†	†	4.73	0.77	9
TBRS Phonological Awareness						
Treatment	†	†	†	5.82	3.57	11
Control	†	†	†	3.33	2.40	9
TBRS Print and Letter Knowledge						
Treatment	†	†	†	3.18	0.93	11
Control	†	†	†	2.80	0.86	9
TBRS Written Expression						
Treatment	†	†	†	2.36	0.62	11
Control	†	†	†	3.22	1.14	9
TBRS Math Concepts						
Treatment	†	†	†	2.36	0.80	11
Control	†	†	†	3.16	1.37	9

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-12a. Unadjusted mean scores of child-level outcome measures, Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: California and New York

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	97.82	13.15	157	98.75	12.55	148	103.21	12.12	145
Control	94.94	16.38	157	95.39	12.31	148	101.67	12.71	139
CMA-A Mathematics Composite									
Treatment	0.40	0.24	157	0.65	0.22	148	0.71	0.16	145
Control	0.40	0.25	157	0.54	0.21	149	0.68	0.17	139
Shape Composition ¹									
Treatment	0.92	0.88	157	2.20	0.85	148	2.63	0.68	145
Control	0.82	0.88	157	1.38	0.87	149	2.34	0.81	139
Reading									
TERA									
Treatment	86.91	13.01	149	90.62	12.64	141	96.76	14.26	136
Control	85.49	11.45	135	89.88	13.28	129	93.91	15.25	129
WJ Letter Word Identification									
Treatment	100.27	17.14	157	102.54	13.67	148	110.81	15.18	145
Control	100.52	15.67	157	100.89	14.34	149	109.06	16.45	139
WJ Spelling									
Treatment	93.13	13.36	157	95.39	12.71	148	104.74	12.24	145
Control	90.63	12.26	157	91.55	12.12	149	104.13	14.01	138
Phonological awareness									
Pre-CTOPPP									
Treatment	7.47	3.16	149	9.77	3.87	141	†	†	†
Control	7.61	3.62	135	9.24	4.35	129	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.06	2.90	139
Control	†	†	†	†	†	†	4.25	3.46	129
Language									
PPVT									
Treatment	89.24	12.67	157	94.99	12.19	148	98.34	10.45	144
Control	90.67	13.86	157	93.92	13.89	148	97.08	11.01	137
TOLD									
Treatment	8.70	2.53	131	9.51	2.37	141	10.09	2.40	138
Control	8.25	2.41	120	9.25	2.29	128	9.82	2.35	129
Behavior									
SSRS Social Skills									
Treatment	104.56	13.63	154	110.81	13.36	150	101.36	14.33	109
Control	106.66	15.02	153	108.65	14.95	147	100.59	15.06	113
SSRS Problem Behavior									
Treatment	97.84	12.80	157	98.28	13.61	149	100.31	11.95	117
Control	96.13	13.13	155	98.26	13.11	145	101.20	13.87	113
PLBS									
Treatment	53.24	9.10	157	53.63	10.39	150	†	†	†
Control	53.71	10.28	152	52.87	10.87	145	†	†	†
LBS									
Treatment	†	†	†	†	†	†	47.87	7.78	119
Control	†	†	†	†	†	†	46.58	10.36	113

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-12b. Unadjusted mean scores of classroom-level outcome measures, Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: California and New York

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	3.41	0.68	19	3.70	0.90	20
Control	3.66	0.91	20	3.55	0.85	20
Teacher-child interaction						
Arnett Detachment						
Treatment	1.72	0.53	19	1.92	0.80	19
Control	1.88	0.87	20	2.23	0.95	20
Arnett Harshness						
Treatment	1.43	0.43	19	1.70	0.69	19
Control	1.42	0.61	20	1.68	0.65	20
Arnett Permissiveness						
Treatment	2.11	0.33	19	2.12	0.34	19
Control	2.12	0.39	20	2.25	0.34	20
Arnett Positive Interactions						
Treatment	2.77	0.57	19	2.84	0.79	19
Control	2.80	0.92	20	2.72	0.88	20
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	1.96	2.43	19
Control	†	†	†	1.56	1.70	20
TBRS Oral Language						
Treatment	†	†	†	3.49	1.40	19
Control	†	†	†	2.81	1.68	20
TBRS Phonological Awareness						
Treatment	†	†	†	3.11	2.38	19
Control	†	†	†	2.30	1.92	20
TBRS Print and Letter Knowledge						
Treatment	†	†	†	1.95	0.72	19
Control	†	†	†	1.90	0.57	20
TBRS Written Expression						
Treatment	†	†	†	2.19	0.72	19
Control	†	†	†	2.27	0.93	20
TBRS Math Concepts						
Treatment	†	†	†	2.65	1.48	19
Control	†	†	†	1.74	0.70	20

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-13a. Unadjusted mean scores of child-level outcome measures, Project Approach: Wisconsin

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	93.15	17.91	114	98.13	13.23	105	100.48	12.48	83
Control	99.01	14.91	89	103.91	13.00	87	103.34	11.43	67
CMA-A Mathematics Composite									
Treatment	0.36	0.23	114	0.60	0.23	105	0.67	0.18	83
Control	0.46	0.25	90	0.67	0.22	87	0.72	0.14	67
Shape Composition ¹									
Treatment	1.11	0.89	114	1.98	0.99	105	2.48	0.70	83
Control	1.37	0.81	90	2.22	0.78	87	2.64	0.60	67
Reading									
TERA									
Treatment	87.31	13.13	114	94.18	16.55	105	96.12	16.87	83
Control	90.48	12.13	89	99.78	13.86	87	99.33	14.46	67
WJ Letter Word Identification									
Treatment	101.18	18.55	114	107.37	14.37	105	108.13	14.80	83
Control	100.69	13.26	90	108.18	12.28	87	111.10	11.05	67
WJ Spelling									
Treatment	92.02	11.70	114	98.79	14.29	105	103.17	13.78	83
Control	94.99	12.34	90	103.64	14.06	87	109.66	11.98	67
Phonological awareness									
Pre-CTOPPP									
Treatment	8.01	3.57	114	10.07	3.70	105	†	†	†
Control	8.31	3.52	90	11.29	4.07	87	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	4.57	3.82	83
Control	†	†	†	†	†	†	5.45	3.34	67
Language									
PPVT									
Treatment	88.82	16.16	114	95.96	13.87	105	98.43	12.40	83
Control	95.07	17.26	90	99.60	15.23	87	103.81	10.06	67
TOLD									
Treatment	8.54	2.38	114	9.41	2.36	105	10.29	2.36	83
Control	9.30	2.68	90	10.22	2.89	87	10.61	2.71	67
Behavior									
SSRS Social Skills									
Treatment	102.18	13.78	109	107.48	13.48	103	95.97	16.65	63
Control	103.84	13.68	76	108.52	12.30	86	102.50	10.98	52
SSRS Problem Behavior									
Treatment	101.69	13.81	112	102.25	13.20	104	105.05	14.69	63
Control	97.31	12.20	81	96.53	11.40	86	95.54	11.66	52
PLBS									
Treatment	51.78	10.16	112	52.01	9.69	104	†	†	†
Control	51.88	10.11	77	56.27	9.11	85	†	†	†
LBS									
Treatment	†	†	†	†	†	†	45.31	12.56	62
Control	†	†	†	†	†	†	49.79	7.17	52

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-13b. Unadjusted mean scores of classroom-level outcome measures, Project Approach: Wisconsin

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	2.94	0.52	7	2.61	0.48	7
Control	2.85	0.70	6	2.73	0.57	6
Teacher-child interaction						
Arnett Detachment						
Treatment	1.46	0.64	7	1.82	0.59	7
Control	1.46	0.51	6	1.79	0.53	6
Arnett Harshness						
Treatment	2.19	0.75	7	2.24	0.90	7
Control	1.74	0.55	6	2.24	0.88	6
Arnett Permissiveness						
Treatment	2.10	0.57	7	1.95	0.23	7
Control	2.17	0.66	6	2.06	0.25	6
Arnett Positive Interactions						
Treatment	2.83	0.74	7	2.41	0.75	7
Control	3.20	0.58	6	2.20	0.79	6
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	3.83	3.44	6
Control	†	†	†	4.05	1.96	6
TBRS Oral Language						
Treatment	†	†	†	3.88	1.16	6
Control	†	†	†	3.79	1.15	6
TBRS Phonological Awareness						
Treatment	†	†	†	2.50	2.35	6
Control	†	†	†	4.00	0.89	6
TBRS Print and Letter Knowledge						
Treatment	†	†	†	2.61	1.05	6
Control	†	†	†	2.06	0.52	6
TBRS Written Expression						
Treatment	†	†	†	2.44	1.31	6
Control	†	†	†	1.89	1.11	6
TBRS Math Concepts						
Treatment	†	†	†	1.93	0.48	6
Control	†	†	†	1.83	0.55	6

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-14a. Unadjusted mean scores of child-level outcome measures, Project Construct: Missouri

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	98.80	13.16	104	99.17	13.50	106	105.24	11.38	99
Control	99.53	15.13	99	98.50	14.26	90	104.01	12.99	80
CMA-A Mathematics Composite									
Treatment	0.57	0.22	120	0.68	0.20	113	0.74	0.15	102
Control	0.59	0.24	108	0.70	0.20	94	0.76	0.16	86
Shape Composition ¹									
Treatment	1.33	0.99	117	1.66	0.90	113	2.52	0.70	102
Control	1.57	0.94	107	2.13	0.89	94	2.49	0.72	86
Reading									
TERA									
Treatment	88.47	12.46	118	86.60	14.27	111	94.12	15.55	99
Control	89.57	12.75	104	87.46	15.26	92	93.75	14.16	83
WJ Letter Word Identification									
Treatment	99.93	13.93	111	99.35	14.01	108	107.40	13.32	99
Control	100.30	13.37	104	100.06	11.48	87	105.34	12.86	83
WJ Spelling									
Treatment	94.47	15.26	112	95.15	14.42	109	107.92	11.36	93
Control	96.76	14.91	103	96.81	13.22	93	107.24	13.12	75
Phonological awareness									
Pre-CTOPPP									
Treatment	8.69	3.96	120	10.51	4.31	113	†	†	†
Control	7.81	4.18	108	10.63	4.50	94	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	5.37	4.09	102
Control	†	†	†	†	†	†	5.19	3.20	86
Language									
PPVT									
Treatment	94.40	14.53	120	96.14	15.47	113	100.75	12.30	96
Control	93.54	15.03	108	95.11	14.92	94	97.98	13.54	85
TOLD									
Treatment	9.60	2.79	119	9.43	2.94	110	11.11	2.39	101
Control	8.94	2.75	107	9.49	2.75	94	11.13	2.45	84
Behavior									
SSRS Social Skills									
Treatment	102.20	13.95	113	108.05	17.28	102	102.70	15.93	79
Control	97.18	13.58	103	102.36	13.46	81	100.47	15.37	75
SSRS Problem Behavior									
Treatment	100.47	13.72	116	99.83	14.04	102	101.83	13.52	81
Control	101.76	12.98	104	103.10	13.14	86	100.83	13.98	76
PLBS									
Treatment	50.82	10.55	119	52.07	11.29	99	†	†	†
Control	49.21	11.93	106	50.28	10.99	86	†	†	†
LBS									
Treatment	†	†	†	†	†	†	47.29	11.35	82
Control	†	†	†	†	†	†	47.91	10.30	75

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-14b. Unadjusted mean scores of classroom-level outcome measures, Project Construct: Missouri

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	3.95	0.96	12	3.71	1.04	12
Control	3.69	0.73	11	3.36	0.76	9
Teacher-child interaction						
Arnett Detachment						
Treatment	1.31	0.40	12	1.27	0.42	12
Control	1.20	0.46	11	1.22	0.34	9
Arnett Harshness						
Treatment	1.32	0.19	12	1.34	0.21	12
Control	1.48	0.53	11	1.33	0.29	9
Arnett Permissiveness						
Treatment	3.50	0.27	12	3.42	0.49	12
Control	3.42	0.52	11	3.44	0.44	9
Arnett Positive Interactions						
Treatment	2.81	0.37	12	2.79	0.57	12
Control	2.72	0.59	11	2.62	0.43	9
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	1.29	1.69	12
Control	†	†	†	0.63	0.85	9
TBRS Oral Language						
Treatment	†	†	†	3.01	1.09	12
Control	†	†	†	2.48	1.07	9
TBRS Phonological Awareness						
Treatment	†	†	†	2.00	1.04	12
Control	†	†	†	2.22	1.30	9
TBRS Print and Letter Knowledge						
Treatment	†	†	†	1.58	0.34	12
Control	†	†	†	1.52	0.39	9
TBRS Written Expression						
Treatment	†	†	†	1.75	0.53	12
Control	†	†	†	1.63	0.59	9
TBRS Math Concepts						
Treatment	†	†	†	1.40	0.95	12
Control	†	†	†	1.10	0.76	9

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table C-15a. Unadjusted mean scores of child-level outcome measures, Ready, Set, Leap!: New Jersey

Measure	Child-level outcome measures								
	Fall Pre-K			Spring Pre-K			Spring K		
	Mean	STD	N	Mean	STD	N	Mean	STD	N
Mathematics									
WJ Applied Problems									
Treatment	92.86	13.41	132	91.92	14.40	136	96.33	13.48	128
Control	90.56	14.87	119	91.32	12.17	118	96.08	10.98	115
CMA-A Mathematics Composite									
Treatment	0.39	0.24	142	0.58	0.22	140	0.67	0.19	132
Control	0.42	0.23	132	0.65	0.24	122	0.70	0.17	117
Shape Composition ¹									
Treatment	1.11	0.88	141	1.72	0.92	140	1.98	0.93	132
Control	0.93	0.86	132	1.69	0.89	122	1.98	0.98	115
Reading									
TERA									
Treatment	90.03	12.37	137	91.47	12.98	139	90.16	13.52	123
Control	89.03	14.00	120	90.39	14.87	118	88.95	13.42	111
WJ Letter Word Identification									
Treatment	107.20	16.00	133	109.83	12.05	139	107.23	12.91	129
Control	107.44	16.28	120	109.12	14.41	118	107.91	12.80	114
WJ Spelling									
Treatment	98.61	16.65	137	102.00	14.57	127	108.69	13.60	117
Control	97.87	17.98	126	98.22	15.92	109	108.03	13.94	106
Phonological awareness									
Pre-CTOPPP									
Treatment	6.43	3.01	140	8.01	3.99	140	†	†	†
Control	6.52	3.56	129	8.39	4.12	122	†	†	†
CTOPP									
Treatment	†	†	†	†	†	†	2.71	2.81	132
Control	†	†	†	†	†	†	2.85	3.40	117
Language									
PPVT									
Treatment	85.10	12.35	139	87.10	14.57	139	90.02	12.35	128
Control	81.43	14.17	128	84.96	15.21	121	90.25	11.78	112
TOLD									
Treatment	8.21	2.72	136	8.33	2.58	139	8.94	2.74	131
Control	7.98	2.41	126	8.74	2.61	119	9.08	2.79	116
Behavior									
SSRS Social Skills									
Treatment	101.41	17.32	140	105.38	15.68	138	96.54	15.12	125
Control	102.57	14.99	123	106.85	15.47	117	97.18	14.41	109
SSRS Problem Behavior									
Treatment	101.67	14.11	141	103.22	12.83	139	103.12	14.70	123
Control	101.39	12.15	127	103.63	12.83	121	101.91	13.54	113
PLBS									
Treatment	51.43	10.11	136	51.56	10.46	139	†	†	†
Control	50.60	9.79	124	51.03	10.76	120	†	†	†
LBS									
Treatment	†	†	†	†	†	†	44.08	12.05	124
Control	†	†	†	†	†	†	44.00	9.55	113

† Not applicable.

¹ Building Blocks, Shape Composition task

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table C-15b. Unadjusted mean scores of classroom-level outcome measures, Ready, Set, Leap!: New Jersey

Measure	Classroom-level outcome measures					
	Fall Pre-K			Spring Pre-K		
	Mean	STD	N	Mean	STD	N
Global classroom quality						
ECERS-R						
Treatment	4.42	0.70	21	4.16	0.94	21
Control	4.45	1.19	18	4.07	1.24	18
Teacher-child interaction						
Arnett Detachment						
Treatment	1.14	0.30	21	1.20	0.28	21
Control	1.10	0.23	18	1.14	0.18	18
Arnett Harshness						
Treatment	1.39	0.39	21	1.44	0.45	21
Control	1.31	0.28	18	1.30	0.20	18
Arnett Permissiveness						
Treatment	3.38	0.47	21	3.30	0.71	21
Control	3.57	0.48	17	3.47	0.35	17
Arnett Positive Interactions						
Treatment	3.08	0.54	21	2.89	0.84	21
Control	3.18	0.62	18	2.93	0.66	18
Teacher instructional practices						
TBRS Book Reading						
Treatment	†	†	†	3.45	2.98	21
Control	†	†	†	4.39	2.91	18
TBRS Oral Language						
Treatment	†	†	†	3.41	1.61	21
Control	†	†	†	3.87	1.68	18
TBRS Phonological Awareness						
Treatment	†	†	†	5.38	2.71	21
Control	†	†	†	5.06	3.44	18
TBRS Print and Letter Knowledge						
Treatment	†	†	†	2.90	1.23	21
Control	†	†	†	3.02	1.61	18
TBRS Written Expression						
Treatment	†	†	†	2.64	1.37	21
Control	†	†	†	2.61	1.25	18
TBRS Math Concepts						
Treatment	†	†	†	1.78	1.23	21
Control	†	†	†	2.00	1.43	18

† Not applicable. The TBRS data were collected at the spring pre-kindergarten time point only.

NOTE: STD: Standard deviation. Refer to the glossary for abbreviations of the measures.

SOURCE: PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

**Appendix D:
Covariate Adjusted Mean Differences and Standard Errors**

Page intentionally left blank.

Table D-1a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Bright Beginnings: Tennessee

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	-0.96	3.45	4.21	3.21	3.33	2.94
CMA-A Mathematics Composite	-0.00	0.04	0.04	0.04	0.02	0.03
Shape Composition ¹	0.07	0.12	-0.03	0.14	0.13	0.12
Reading						
TERA	0.81	0.86	3.43*	1.30	-0.64	1.50
WJ Letter Word Identification	6.50	4.09	8.85	4.34	2.16	4.71
WJ Spelling	0.55	3.56	4.81	3.60	1.55	3.47
Phonological awareness						
Pre-CTOPPP/CTOPP	0.20	0.63	-0.36	0.55	0.04	0.41
Language						
PPVT	-0.96	2.20	2.38	2.29	1.22	2.02
TOLD	-0.55	0.62	0.49	0.69	0.82	0.56
Behavior						
SSRS Social Skills	-5.84	8.21	-2.09	6.06	-0.55	4.50
SSRS Problem Behavior	1.01	5.15	2.63	4.23	3.28	2.49
PLBS/LBS	0.58	5.57	0.32	3.30	-3.69	2.41

* p < .05

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-1b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Bright Beginnings: Tennessee*

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	0.98*	0.37	0.56	0.41
Teacher-child interaction				
Arnett Detachment	-0.76*	0.33	0.12	0.36
Arnett Harshness	-0.23	0.19	0.04	0.21
Arnett Permissiveness	-0.16	0.19	0.06	0.21
Arnett Positive Interactions	0.61	0.36	0.29	0.40
Teacher instructional practices				
TBRS Book Reading	†	†	2.40	1.49
TBRS Oral Language	†	†	0.49	0.78
TBRS Phonological Awareness	†	†	4.54*	1.73
TBRS Print and Letter Knowledge	†	†	1.54*	0.62
TBRS Written Expression	†	†	1.66*	0.61
TBRS Math Concepts	†	†	1.09	0.57

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-2a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Creative Curriculum: Tennessee

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	2.38	3.47	3.97	3.22	3.91	2.97
CMA-A Mathematics Composite	0.02	0.04	0.02	0.04	0.01	0.03
Shape Composition ¹	0.08	0.12	-0.11	0.14	-0.00	0.12
Reading						
TERA	-0.47	0.86	0.14	1.29	0.74	1.52
WJ Letter Word Identification	6.52	4.11	3.59	4.35	8.39	4.75
WJ Spelling	-1.44	3.58	4.45	3.62	5.98	3.51
Phonological awareness						
Pre-CTOPPP/CTOPP	-0.61	0.64	-0.04	0.56	0.20	0.41
Language						
PPVT	-0.22	2.21	4.13	2.28	2.08	2.03
TOLD	-0.22	0.62	0.33	0.69	0.54	0.57
Behavior						
SSRS Social Skills	-4.29	8.22	1.83	6.07	5.55	4.57
SSRS Problem Behavior	0.26	5.15	0.68	4.24	-0.70	2.53
PLBS/LBS	0.04	5.57	1.73	3.30	0.89	2.45

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-2b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Creative Curriculum: Tennessee*

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	1.35	0.38	0.32	0.40
Teacher-child interaction				
Arnett Detachment	-0.72	0.34	-0.12	0.40
Arnett Harshness	-0.13	0.20	-0.05	0.24
Arnett Permissiveness	-0.04	0.20	0.18	0.23
Arnett Positive Interactions	0.55	0.38	-0.11	0.44
Teacher instructional practices				
TBRS Book Reading	†	†	-0.97	1.65
TBRS Oral Language	†	†	-0.10	0.86
TBRS Phonological Awareness	†	†	3.03	1.92
TBRS Print and Letter Knowledge	†	†	0.82	0.69
TBRS Written Expression	†	†	1.40	0.68
TBRS Math Concepts	†	†	1.06	0.68

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-3a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Creative Curriculum: North Carolina and Georgia

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	3.98	3.98	4.62	3.59	2.12	3.37
CMA-A Mathematics Composite	0.00	0.03	-0.02	0.04	0.03	0.04
Shape Composition ¹	0.01	0.14	0.17	0.15	-0.00	0.15
Reading						
TERA	0.96	1.12	-0.58	1.43	-0.30	1.82
WJ Letter Word Identification	-4.71	6.34	-1.87	6.08	-0.07	6.77
WJ Spelling	0.32	4.79	-4.56	5.06	-1.29	5.17
Phonological awareness						
Pre-CTOPPP/CTOPP	0.01	0.52	0.19	0.71	0.17	0.51
Language						
PPVT	0.25	2.09	1.22	2.26	2.35	2.41
TOLD	0.65	0.63	-0.74	0.73	-0.82	0.69
Behavior						
SSRS Social Skills	2.69	4.39	0.01	3.61	-1.60	3.23
SSRS Problem Behavior	-1.53	4.04	-1.77	3.47	1.01	3.80
PLBS/LBS	1.87	2.73	0.17	2.61	-2.33	2.09

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-3b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Creative Curriculum*: North Carolina and Georgia

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	0.37	0.38	1.06 *	0.40
Teacher-child interaction				
Arnett Detachment	-0.40	0.37	-1.06 *	0.38
Arnett Harshness	-0.47	0.35	-0.49	0.36
Arnett Permissiveness	0.20	0.30	-0.57	0.30
Arnett Positive Interactions	0.27	0.33	1.02 **	0.33
Teacher instructional practices				
TBRS Book Reading	†	†	0.39	1.22
TBRS Oral Language	†	†	1.48 **	0.43
TBRS Phonological Awareness	†	†	-0.32	1.68
TBRS Print and Letter Knowledge	†	†	0.72	0.43
TBRS Written Expression	†	†	1.61 **	0.49
TBRS Math Concepts	†	†	0.84	0.60

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$; ** $p < .01$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-4a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Creative Curriculum with Ladders to Literacy: New Hampshire

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	-3.17	4.49	-3.01	5.69	-7.00	4.93
CMA-A Mathematics Composite	0.03	0.04	0.04	0.06	-0.05	0.04
Shape Composition ¹	-0.06	0.16	0.02	0.24	-0.09	0.18
Reading						
TERA	0.69	1.71	-2.37	2.39	-4.23	2.50
WJ Letter Word Identification	-2.22	5.60	-4.09	6.92	-6.79	6.67
WJ Spelling	-5.86	4.90	7.35	6.16	-1.89	6.17
Phonological awareness						
Pre-CTOPPP/CTOPP	-0.14	1.41	-0.55	0.90	-0.41	0.98
Language						
PPVT	-0.85	4.75	-7.19	5.28	-5.64	5.05
TOLD	-0.03	1.07	-1.07	1.29	-0.29	1.10
Behavior						
SSRS Social Skills	-4.62	7.04	-1.00	2.79	2.23	5.09
SSRS Problem Behavior	-0.33	4.06	-0.33	3.01	0.25	4.43
PLBS/LBS	-1.88	2.62	-0.30	1.65	-1.16	3.08

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-4b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Creative Curriculum with Ladders to Literacy*: New Hampshire

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	0.45	0.32	-0.56	0.46
Teacher-child interaction				
Arnett Detachment	-0.18	0.34	0.39	0.52
Arnett Harshness	0.22	0.18	-0.09	0.27
Arnett Permissiveness	0.10	0.19	0.35	0.30
Arnett Positive Interactions	0.47	0.29	0.03	0.45
Teacher instructional practices				
TBRS Book Reading	†	†	-0.74	1.47
TBRS Oral Language	†	†	-0.59	0.67
TBRS Phonological Awareness	†	†	-0.60	1.67
TBRS Print and Letter Knowledge	†	†	0.71	0.50
TBRS Written Expression	†	†	1.24*	0.38
TBRS Math Concepts	†	†	0.63	0.83

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-5a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Curiosity Corner: Florida, Kansas, and New Jersey

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	1.26	3.34	1.90	3.14	5.11	3.28
CMA-A Mathematics Composite	0.00	0.05	0.00	0.05	-0.01	0.04
Shape Composition ¹	-0.06	0.20	0.15	0.20	0.31	0.18
Reading						
TERA	2.33	1.23	0.83	1.40	3.50	1.57
WJ Letter Word Identification	5.99	3.70	2.42	3.96	11.26*	4.43
WJ Spelling	-3.28	5.26	0.97	4.66	5.60	4.59
Phonological awareness						
Pre-CTOPPP/CTOPP	-0.05	0.46	-0.04	0.75	0.94	0.81
Language						
PPVT	-0.67	2.80	-0.17	2.77	2.42	2.69
TOLD	-0.17	0.85	-0.38	0.79	0.74	0.65
Behavior						
SSRS Social Skills	-3.59	3.58	-1.40	2.49	4.87	3.77
SSRS Problem Behavior	6.63*	2.69	0.84	1.98	-1.15	3.09
PLBS/LBS	-4.01	2.20	0.24	1.94	1.11	2.38

* p < .05

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-5b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Curiosity Corner*: Florida, Kansas, and New Jersey

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	-0.70	0.43	-0.52	0.44
Teacher-child interaction				
Arnett Detachment	-0.12	0.36	-0.36	0.38
Arnett Harshness	0.12	0.25	0.06	0.24
Arnett Permissiveness	-0.70 *	0.27	-0.47	0.26
Arnett Positive Interactions	0.27	0.25	0.01	0.24
Teacher instructional practices				
TBRS Book Reading	†	†	4.46 ***	1.10
TBRS Oral Language	†	†	0.40	0.73
TBRS Phonological Awareness	†	†	1.08	1.70
TBRS Print and Letter Knowledge	†	†	-0.92	0.59
TBRS Written Expression	†	†	-0.44	0.51
TBRS Math Concepts	†	†	-0.34	0.65

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$; *** $p < .001$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-6a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Doors to Discovery: Texas

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	2.50	4.27	0.25	3.62	-0.56	3.58
CMA-A Mathematics Composite	0.01	0.04	0.03	0.03	-0.04	0.03
Shape Composition ¹	-0.10	0.16	-0.11	0.14	-0.11	0.14
Reading						
TERA	1.79	1.90	0.64	2.07	-0.58	2.26
WJ Letter Word Identification	2.91	4.87	2.78	4.78	-2.54	5.67
WJ Spelling	3.61	4.51	1.54	4.19	-2.95	4.61
Phonological awareness						
Pre-CTOPPP/CTOPP	0.73	0.85	0.67	0.66	-0.36	0.67
Language						
PPVT	4.84	4.95	3.30	4.94	4.13	5.01
TOLD	1.92*	0.86	0.86	0.80	0.33	0.84
Behavior						
SSRS Social Skills	-0.09	3.99	-2.75	2.86	-0.73	3.79
SSRS Problem Behavior	-4.78	3.26	1.54	2.05	5.80	3.04
PLBS/LBS	0.57	2.51	-2.42	2.15	-2.49	1.80

* p < .05

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D6-b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Doors to Discovery: Texas*

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	0.16	0.23	0.25	0.23
Teacher-child interaction				
Arnett Detachment	-0.24	0.24	-0.05	0.24
Arnett Harshness	-0.24	0.22	-0.26	0.23
Arnett Permissiveness	0.33*	0.14	0.04	0.14
Arnett Positive Interactions	0.31	0.26	0.30	0.26
Teacher instructional practices				
TBRS Book Reading	†	†	2.95**	1.04
TBRS Oral Language	†	†	0.92	0.55
TBRS Phonological Awareness	†	†	2.06	1.49
TBRS Print and Letter Knowledge	†	†	1.03*	0.51
TBRS Written Expression	†	†	0.70	0.52
TBRS Math Concepts	†	†	0.41	0.46

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$; ** $p < .01$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-7a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Let's Begin with the Letter People: Texas

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	-2.09	4.25	-2.53	3.59	-3.22	3.56
CMA-A Mathematics Composite	0.04	0.04	0.04	0.03	-0.02	0.03
Shape Composition ¹	-0.27	0.16	0.20	0.14	-0.06	0.14
Reading						
TERA	-0.26	1.88	0.18	2.06	-1.31	2.25
WJ Letter Word Identification	-2.84	4.84	2.68	4.75	-4.78	5.66
WJ Spelling	0.13	4.48	3.97	4.16	-1.55	4.60
Phonological awareness						
Pre-CTOPPP/CTOPP	0.36	0.85	-0.76	0.66	-0.52	0.67
Language						
PPVT	2.49	4.89	-0.61	4.88	-0.05	4.95
TOLD	0.34	0.85	0.41	0.80	-0.61	0.84
Behavior						
SSRS Social Skills	-5.92	3.93	0.30	2.83	3.53	3.81
SSRS Problem Behavior	-1.62	3.22	0.22	2.01	0.78	3.01
PLBS/LBS	-2.21	2.48	-3.63	2.14	-0.69	1.83

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-7b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Let's Begin with the Letter People*: Texas

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality ECERS-R	0.07	0.21	0.55*	0.22
Teacher-child interaction				
Arnett Detachment	0.12	0.22	-0.05	0.24
Arnett Harshness	-0.10	0.21	-0.56*	0.22
Arnett Permissiveness	0.34**	0.13	-0.02	0.13
Arnett Positive Interactions	0.08	0.24	0.36	0.26
Teacher instructional practices				
TBRS Book Reading	†	†	1.57	0.99
TBRS Oral Language	†	†	0.63	0.52
TBRS Phonological Awareness	†	†	2.59	1.42
TBRS Print and Letter Knowledge	†	†	1.21*	0.48
TBRS Written Expression	†	†	0.78	0.50
TBRS Math Concepts	†	†	0.26	0.44

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$; ** $p < .01$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-8a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Early Literacy and Learning Model: Florida—University of North Florida

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	1.26	3.34	1.90	3.14	5.11	3.28
CMA-A Mathematics Composite	0.00	0.05	0.00	0.05	-0.01	0.04
Shape Composition ¹	0.10	0.14	-0.11	0.15	0.02	0.15
Reading						
TERA	1.06	1.71	1.25	1.81	2.49	1.91
WJ Letter Word Identification	-3.04	5.47	-1.17	5.27	0.07	5.58
WJ Spelling	-5.71	5.06	3.04	5.20	1.20	5.20
Phonological awareness						
Pre-CTOPPP/CTOPP	-0.05	0.46	0.50	0.58	0.27	0.55
Language						
PPVT	3.93	2.35	2.61	2.34	5.15*	2.51
TOLD	0.08	0.64	0.71	0.62	2.01**	0.61
Behavior						
SSRS Social Skills	-3.59	3.58	1.32	2.60	4.82	3.70
SSRS Problem Behavior	-4.57	2.27	-0.16	2.54	3.53	3.26
PLBS/LBS	0.71	1.57	0.76	1.99	0.45	2.46

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

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-8b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Early Literacy and Learning Model: Florida—University of North Florida*

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality ECERS-R	-0.70	0.43	-0.52	0.44
Teacher-child interaction				
Arnett Detachment	-0.12	0.36	-0.36	0.38
Arnett Harshness	-0.01	0.36	-0.35	0.38
Arnett Permissiveness	0.01	0.18	-0.10	0.19
Arnett Positive Interactions	-0.29	0.35	0.22	0.37
Teacher instructional practices				
TBRS Book Reading	†	†	0.77	1.17
TBRS Oral Language	†	†	0.20	0.57
TBRS Phonological Awareness	†	†	1.15	1.02
TBRS Print and Letter Knowledge	†	†	0.39	0.47
TBRS Written Expression	†	†	-0.18	0.36
TBRS Math Concepts	†	†	-0.67	0.36

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-9a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Language-Focused Curriculum: Virginia

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	-1.76	2.66	3.26	2.31	1.82	2.11
CMA-A Mathematics Composite	-0.00	0.04	0.02	0.04	0.00	0.03
Shape Composition ¹	0.06	0.12	0.07	0.12	0.05	0.12
Reading						
TERA	0.47	1.16	1.16	1.37	0.35	1.44
WJ Letter Word Identification	2.05	3.53	2.57	3.38	0.43	3.34
WJ Spelling	2.65	4.60	6.36	4.83	2.91	4.29
Phonological awareness						
Pre-CTOPPP/CTOPP	-0.23	0.75	1.08	0.72	0.09	0.49
Language						
PPVT	-1.85	2.69	0.34	2.49	-1.37	2.35
TOLD	-0.45	0.76	0.03	0.69	-0.30	0.60
Behavior						
SSRS Social Skills	1.33	4.50	-8.16	3.84	-1.04	3.11
SSRS Problem Behavior	5.13	3.72	2.77	2.66	-0.72	2.72
PLBS/LBS	-2.69	2.09	-2.41	2.19	1.07	2.40

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-9b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Language-Focused Curriculum: Virginia*

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	—	—	0.34	0.43
Teacher-child interaction				
Arnett Detachment	—	—	-0.59	0.44
Arnett Harshness	—	—	-0.19	0.09
Arnett Permissiveness	—	—	0.07	0.12
Arnett Positive Interactions	—	—	0.27	0.35
Teacher instructional practices				
TBRS Book Reading	†	†	-1.51	1.10
TBRS Oral Language	†	†	0.90	0.73
TBRS Phonological Awareness	†	†	1.65	1.18
TBRS Print and Letter Knowledge	†	†	0.30	0.28
TBRS Written Expression	†	†	0.75	0.55
TBRS Math Concepts	†	†	0.13	0.41

— Not available. Data were collected but not reported.

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

NOTE: SE: Standard error. For the *Language-Focused Curriculum*, we did not conduct analyses using the ECERS and Arnett data because of unreliable data. During the baseline data collection, one observer completed the observational ratings in 8 of the 12 classrooms at this research site. It was later determined that the ECERS-R and Arnett ratings from these eight classrooms were inflated. Due to concerns with the integrity of the data from these eight classrooms, the decision was made to exclude the classroom quality and teacher-child relationships data for this site from the report. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-10a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Literacy Express: Florida—Florida State University

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	0.62	3.77	1.26	3.22	-0.44	3.36
CMA-A Mathematics Composite	-0.04	0.04	-0.01	0.03	-0.05	0.03
Shape Composition ¹	-0.12	0.17	-0.01	0.15	-0.13	0.16
Reading						
TERA	-0.89	1.26	1.26	1.43	-0.85	1.82
WJ Letter Word Identification	11.23*	4.68	7.70	4.24	2.16	5.08
WJ Spelling	3.18	5.15	1.36	4.56	1.55	4.99
Phonological awareness						
Pre-CTOPPP/CTOPP	-0.78	0.62	0.63	0.53	0.27	0.50
Language						
PPVT	1.25	2.80	2.99	2.54	2.86	2.75
TOLD	-1.15	0.92	-0.22	0.73	0.52	0.79
Behavior						
SSRS Social Skills	5.33	3.76	-3.18	3.33	-5.52	3.01
SSRS Problem Behavior	-6.94	3.52	-0.03	3.05	3.00	2.34
PLBS/LBS	4.03	3.27	-0.26	2.09	-3.74	1.68

* p < .05

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-10b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Literacy Express: Florida—Florida State University*

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	0.09	0.34	0.89 *	0.34
Teacher-child interaction				
Arnett Detachment	0.05	0.21	-0.35	0.21
Arnett Harshness	-0.04	0.13	-0.24	0.12
Arnett Permissiveness	-0.32	0.20	0.20	0.19
Arnett Positive Interactions	-0.27	0.26	0.25	0.25
Teacher instructional practices				
TBRS Book Reading	†	†	0.85	0.92
TBRS Oral Language	†	†	0.21	0.42
TBRS Phonological Awareness	†	†	3.43 *	1.46
TBRS Print and Letter Knowledge	†	†	0.88	0.44
TBRS Written Expression	†	†	-0.03	0.45
TBRS Math Concepts	†	†	-0.15	0.56

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-11a. Covariate adjusted mean differences and standard errors of child-level outcome measures, DLM Early Childhood Express supplemented with Open Court Reading Pre-K: Florida—Florida State University

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	4.76	3.59	8.26 **	3.13	10.84 ***	3.25
CMA-A Mathematics Composite	-0.02	0.04	0.04	0.03	0.03	0.03
Shape Composition ¹	-0.08	0.16	0.23	0.15	0.09	0.15
Reading						
TERA	2.12	1.22	5.42 ***	1.40	6.04 **	1.79
WJ Letter Word Identification	11.06 *	4.48	13.65 **	4.12	13.30 **	4.93
WJ Spelling	7.10	4.92	12.69 **	4.45	6.12	4.85
Phonological awareness						
Pre-CTOPPP/CTOPP	0.32	0.59	1.02	0.52	1.23 *	0.49
Language						
PPVT	5.38	2.68	6.64 *	2.47	7.89 **	2.68
TOLD	1.98 *	0.87	2.09 **	0.71	2.40 **	0.78
Behavior						
SSRS Social Skills	4.82	3.60	-4.22	3.19	-2.84	2.82
SSRS Problem Behavior	-0.21	3.42	1.64	2.97	0.18	2.25
PLBS/LBS	-1.08	3.19	-0.96	2.03	-1.26	1.64

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

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-11b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, DLM Early Childhood Express supplemented with Open Court Reading Pre-K: Florida—Florida State University

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	-0.08	0.35	0.24	0.35
Teacher-child interaction				
Arnett Detachment	0.02	0.22	-0.03	0.21
Arnett Harshness	-0.12	0.13	-0.20	0.13
Arnett Permissiveness	-0.12	0.20	0.02	0.20
Arnett Positive Interactions	-0.23	0.26	0.22	0.26
Teacher instructional practices				
TBRS Book Reading	†	†	0.01	1.04
TBRS Oral Language	†	†	-0.25	0.48
TBRS Phonological Awareness	†	†	4.37*	1.66
TBRS Print and Letter Knowledge	†	†	0.82	0.50
TBRS Written Expression	†	†	-0.52	0.51
TBRS Math Concepts	†	†	-0.51	0.64

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

* $p < .05$

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-12a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: California and New York

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	3.77	3.11	4.23	2.94	2.64	2.87
CMA-A Mathematics Composite	0.02	0.03	0.10**	0.03	0.03	0.03
Shape Composition ¹	0.21*	0.11	0.83****	0.11	0.35***	0.10
Reading						
TERA	1.07	1.36	0.99	1.51	2.40	1.58
WJ Letter Word Identification	-3.74	5.11	-0.31	5.20	5.40	5.95
WJ Spelling	3.68	4.23	4.77	4.25	0.82	4.27
Phonological awareness						
Pre-CTOPPP/CTOPP	-0.48	0.48	0.43	0.56	-0.37	0.40
Language						
PPVT	0.81	2.59	2.95	2.75	1.80	2.63
TOLD	0.59	0.59	0.77	0.62	0.37	0.53
Behavior						
SSRS Social Skills	-1.38	2.85	3.41	2.50	0.90	2.23
SSRS Problem Behavior	0.89	2.59	-1.41	2.20	-0.15	2.07
PLBS/LBS	-0.67	1.85	0.84	1.68	0.05	1.44

* p < .05; ** p < .01; *** p < .001; **** p < .0001

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-12b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: California and New York

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	-0.39	0.24	0.05	0.26
Teacher-child interaction				
Arnett Detachment	0.02	0.24	-0.29	0.26
Arnett Harshness	0.10	0.18	0.11	0.20
Arnett Permissiveness	0.01	0.11	-0.15	0.12
Arnett Positive Interactions	-0.15	0.23	0.13	0.25
Teacher instructional practices				
TBRS Book Reading	†	†	0.14	0.63
TBRS Oral Language	†	†	0.30	0.47
TBRS Phonological Awareness	†	†	0.83	0.70
TBRS Print and Letter Knowledge	†	†	0.05	0.23
TBRS Written Expression	†	†	-0.09	0.28
TBRS Math Concepts	†	†	0.66	0.37

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-13a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Project Approach: Wisconsin

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	-1.17	5.27	1.62	4.71	6.03	4.71
CMA-A Mathematics Composite	-0.02	0.05	0.04	0.05	0.05	0.05
Shape Composition ¹	0.02	0.15	0.24	0.16	0.21	0.15
Reading						
TERA	3.42	2.93	1.31	3.03	2.72	3.12
WJ Letter Word Identification	12.75	8.79	11.24	8.55	0.70	9.05
WJ Spelling	7.71	8.88	6.35	8.80	3.33	8.76
Phonological awareness						
Pre-CTOPPP/CTOPP	0.48	0.60	-0.99	0.50	-0.63	0.81
Language						
PPVT	-0.37	3.84	3.10	3.70	1.85	3.69
TOLD	-0.12	0.86	0.75	0.83	1.55	0.82
Behavior						
SSRS Social Skills	-0.68	3.07	0.79	2.65	-6.38*	2.71
SSRS Problem Behavior	4.83	3.14	3.04	2.92	6.70*	2.69
PLBS/LBS	-0.11	2.80	-3.68*	1.42	-4.48	2.25

* p < .05

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-13b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, Project Approach: Wisconsin

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	0.47	0.45	-0.11	0.36
Teacher-child interaction				
Arnett Detachment	0.12	0.48	0.33	0.39
Arnett Harshness	0.83	0.60	0.68	0.48
Arnett Permissiveness	-0.14	0.37	-0.20	0.30
Arnett Positive Interactions	-0.71	0.60	-0.72	0.48
Teacher instructional practices				
TBRS Book Reading	†	†	-2.14	2.23
TBRS Oral Language	†	†	-0.49	0.96
TBRS Phonological Awareness	†	†	-2.12	1.45
TBRS Print and Letter Knowledge	†	†	0.29	0.62
TBRS Written Expression	†	†	0.76	1.06
TBRS Math Concept	†	†	-0.33	0.38

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-14a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Project Construct: Missouri

Outcome measures	Child- level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	-0.93	3.39	1.17	3.00	1.59	2.90
CMA-A Mathematics Composite	-0.01	0.03	-0.02	0.03	-0.01	0.03
Shape Composition ¹	-0.15	0.16	-0.40**	0.13	0.11	0.13
Reading						
TERA	-0.53	1.38	-0.01	1.45	-0.28	1.59
WJ Letter Word Identification	-0.20	5.25	-1.11	5.15	3.94	6.06
WJ Spelling	-7.84	4.92	-4.22	4.26	-0.08	4.18
Phonological awareness						
Pre-CTOPPP/CTOPP	1.37	0.81	-0.31	0.60	-0.44	0.69
Language						
PPVT	0.36	2.82	0.60	2.77	1.76	2.64
TOLD	0.89	0.79	-0.25	0.68	0.04	0.53
Behavior						
SSRS Social Skills	3.26	3.20	1.05	3.81	1.81	2.86
SSRS Problem Behavior	-2.61	3.51	0.98	3.39	1.03	2.16
PLBS/LBS	1.83	2.72	-1.21	2.73	-0.20	1.68

** p < .01

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-14b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, Project Construct: Missouri

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality				
ECERS-R	0.29	0.39	0.48	0.37
Teacher-child interaction				
Arnett Detachment	0.10	0.20	0.05	0.18
Arnett Harshness	-0.26	0.16	-0.04	0.15
Arnett Permissiveness	0.15	0.21	-0.01	0.20
Arnett Positive Interactions	0.21	0.23	0.23	0.21
Teacher instructional practices				
TBRS Book Reading	†	†	1.14	0.69
TBRS Oral Language	†	†	0.56	0.62
TBRS Phonological Awareness	†	†	0.02	0.61
TBRS Print and Letter Knowledge	†	†	0.12	0.16
TBRS Written Expression	†	†	0.24	0.28
TBRS Math Concepts	†	†	0.46	0.45

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).

Table D-15a. Covariate adjusted mean differences and standard errors of child-level outcome measures, Ready, Set, Leap!: New Jersey

Outcome measures	Child-level outcome measures					
	Fall Pre-K		Spring Pre-K		Spring K	
	Mean difference	SE	Mean difference	SE	Mean difference	SE
Mathematics						
WJ Applied Problems	2.54	2.82	0.88	2.39	-0.01	2.28
CMA-A Mathematics Composite	-0.02	0.03	-0.05*	0.03	-0.02	0.02
Shape Composition ¹	0.23*	0.10	0.07	0.11	0.03	0.12
Reading						
TERA	0.60	1.08	0.71	1.18	0.11	1.32
WJ Letter Word Identification	-0.61	3.64	0.35	3.48	-3.17	4.07
WJ Spelling	1.40	4.06	6.22	3.53	1.29	3.01
Phonological awareness						
Pre-CTOPPP/CTOPP	0.03	0.41	-0.22	0.60	-0.05	0.44
Language						
PPVT	2.99	1.84	2.45	2.05	-0.27	1.79
TOLD	0.18	0.62	-0.54	0.59	-0.13	0.58
Behavior						
SSRS Social Skills	-0.26	3.18	-0.76	2.41	-0.43	2.29
SSRS Problem Behavior	0.16	2.03	-0.01	1.94	1.02	1.97
PLBS/LBS	0.97	1.78	0.13	1.52	-0.15	1.28

* p < .05

¹ Building Blocks, Shape Composition task

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Child Assessment Battery and Teacher Rating Scales of Children's Behavior (Fall 2003, Spring 2004, and Spring 2005).

Table D-15b. Covariate adjusted mean differences and standard errors of classroom-level outcome measures, *Ready, Set, Leap!*: New Jersey

Outcome measures	Classroom-level outcome measures			
	Fall Pre-K		Spring Pre-K	
	Mean difference	SE	Mean difference	SE
Global classroom quality ECERS-R	0.01	0.32	0.16	0.37
Teacher-child interaction				
Arnett Detachment	0.05	0.08	0.05	0.09
Arnett Harshness	0.09	0.11	0.11	0.12
Arnett Permissiveness	-0.18	0.17	-0.13	0.19
Arnett Positive Interactions	-0.04	0.21	0.03	0.24
Teacher instructional practices				
TBRS Book Reading	†	†	-0.54	0.89
TBRS Oral Language	†	†	-0.40	0.56
TBRS Phonological Awareness	†	†	0.69	1.01
TBRS Print and Letter Knowledge	†	†	-0.02	0.50
TBRS Written Expression	†	†	0.13	0.47
TBRS Math Concepts	†	†	-0.14	0.39

† Not applicable. The TBRS data were collected at spring pre-kindergarten time point only.

NOTE: SE: Standard error. Refer to the glossary for abbreviations of the measures.

SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study; PCER Classroom Observation Battery (Fall 2003 and Spring 2004).