

# **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**

<b>Curriculum and publisher</b>	<b>Research team</b>	<b>Research site</b>	<b>Evaluator</b>
<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**

<b>Intervention curriculum</b>	<b>Research site</b>	<b>Control curriculum</b>
<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 <sup>1</sup>	14 classrooms <sup>1</sup>	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 <sup>1</sup>	11 Pre-K programs <sup>1</sup>	T: 123 C: 108
UC-Berkeley	<i>Ready, Set, Leap!</i>	18 classrooms	21 classrooms	T: 149 C: 137

<sup>1</sup> 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, 3<sup>rd</sup> 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**

<b>Outcome</b>	<b>Measure</b>	<b>Times collected</b>	<b>Instrument</b>
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 <sup>1</sup>	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 <sup>2</sup>	Pre-K: fall/spring, K: spring	
Pre-kindergarten behavior <sup>1</sup>	SSRS Social Skills	Pre-K: fall/spring	Teacher report
	SSRS Problem Behavior	Pre-K: fall/spring	
	PLBS	Pre-K: fall/spring	
Kindergarten behavior <sup>1</sup>	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

<sup>1</sup>Pre-kindergarten and kindergarten measures are not on the same scale.

<sup>2</sup>Building 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 <sup>1</sup>	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.

<sup>1</sup> 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, 3<sup>rd</sup> 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: 3<sup>rd</sup> 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.**

Grant sites	July			August				September				October				November			
	14	21	28	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17
Missouri			RA				C								BL				
Florida		RA						C							BL				
New Jersey									RA		C	RA		C					BL
Kansas SFA							RA			C					BL				
Florida SFA							RA			C					BL				
New Jersey SFA								RA		C	C				BL				
Virginia			RA							C					BL				

**Key to Symbols**

<b>RA</b>	Random assignment		Child assessment training
<b>C</b>	Consent gathering		Beginning of implementation
<b>BL</b>	Baseline child assessment		

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

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 <sup>1</sup>	91	92	†
University of North Florida <sup>1</sup>	86	88	†
University of North Florida <sup>1</sup>	86	88	†
UC-Berkeley and University at Buffalo, SUNY <sup>1</sup>	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 <sup>1</sup>	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.

<sup>1</sup> 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**

<b>Research team</b>	<b>Response rate Fall 2003</b>	<b>Percent of sample with data Spring 2004</b>	<b>Percent of sample with data Spring 2005</b>
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**

<b>Research team</b>	<b>Response rate</b>	<b>Percent of sample with data</b>	<b>Percent of sample with data</b>
	<b>Fall 2003</b>	<b>Spring 2004</b>	<b>Spring 2005</b>
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**

<b>Characteristics</b>	<b>Preschool</b>	<b>Kindergarten</b>
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**

<b>Characteristics</b>	<b>Percent or average</b>
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**

<b>Outcome</b>	<b>Measure</b>	<b>Times observed</b>	<b>Model</b>
Reading	TERA	3	Spline repeated measures
	WJ Letter Word Identification	3	Spline repeated measures
	WJ Spelling	3	Spline repeated measures
Phonological awareness <sup>1</sup>	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 <sup>2</sup>	3	Spline repeated measures
Pre-kindergarten behavior <sup>1</sup>	SSRS Social Skills	2	Repeated measures
	SSRS Problem Behavior	2	Repeated measures
	PLBS	2	Repeated measures
Kindergarten behavior <sup>1</sup>	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

<sup>1</sup> Pre-kindergarten and kindergarten measures not on the same scale

<sup>2</sup> 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 ( $\alpha$ ) 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 ( $\rho$ ) 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 <sup>1</sup>	-.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

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

<sup>1</sup> 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 <sup>1</sup>	.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

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ <sup>1</sup> 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 .15R<sup>2</sup>: 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.